

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2003年 1月20日

出 願 番 号

Application Number:

特願2003-011592

[ST.10/C]:

[JP2003-011592]

出 願 人

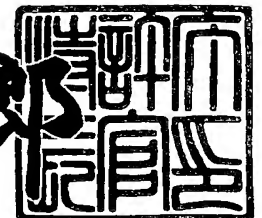
Applicant(s):

株式会社日立製作所

2003年 6月13日

特 許 庁 長 官
Commissioner,
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3046496

【書類名】 特許願

【整理番号】 HI020924

【提出日】 平成15年 1月20日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/06

【発明者】

 【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内

 【氏名】 小林 直孝

【発明者】

 【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内

 【氏名】 ▲高▼田 豊

【発明者】

 【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内

 【氏名】 中山 信一

【発明者】

 【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所ソフトウェア事業部内

 【氏名】 志川 甚一

【発明者】

 【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所ソフトウェア事業部内

 【氏名】 雑賀 信之

【特許出願人】

 【識別番号】 000005108

 【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100071283

【弁理士】

【氏名又は名称】 一色 健輔

【選任した代理人】

【識別番号】 100084906

【弁理士】

【氏名又は名称】 原島 典孝

【選任した代理人】

【識別番号】 100098523

【弁理士】

【氏名又は名称】 黒川 恵

【選任した代理人】

【識別番号】 100112748

【弁理士】

【氏名又は名称】 吉田 浩二

【選任した代理人】

【識別番号】 100110009

【弁理士】

【氏名又は名称】 青木 康

【手数料の表示】

【予納台帳番号】 011785

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 記憶デバイス制御装置の制御方法、及び記憶デバイス制御装置

【特許請求の範囲】

【請求項1】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有するチャンネル制御部を備える記憶デバイス制御装置であって、

前記チャンネル制御部が、あるファイルに対する前記データ入出力要求を前記情報処理装置から受信した場合に、そのファイルに対する排他制御を実行する排他制御手段

を備えることを特徴とする記憶デバイス制御装置。

【請求項2】 前記データ入出力要求は少なくとも2種以上のネットワークファイルシステムプロトコルに従って送信され、あるネットワークファイルシステムプロトコルに従って送信されてくる前記データ入出力要求を受け付けて前記排他制御を実行中に、他のネットワークファイルシステムプロトコルに従って送信されてくる前記データ入出力要求を受け付けた場合に、当該データ入出力要求についても前記排他制御の効果を反映させることを特徴とする請求項1に記載の記憶デバイス制御装置。

【請求項3】 前記記憶デバイスの記憶領域は、この記憶領域上に論理的に設定された論理ボリュームを単位として管理され、

前記I/Oプロセッサは、前記ファイルに対する排他制御に対応して前記論理ボリュームに対する排他制御を実行することを特徴とする請求項1に記載の記憶デバイス制御装置。

【請求項4】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有するチャンネル制御部を備える記憶デバイス制御装置であって、

前記情報処理装置から、あるファイルについての前記記憶デバイスの記憶領域上の記憶位置を特定する情報のリクエストを受信して、前記情報を前記情報処理装置に送信する手段と、

前記情報処理装置から前記情報に基づいて生成されたブロック単位でのデータ読み出し要求を受信して、これに対応する I / O 要求を前記記憶デバイスに対して出力する手段と、

前記記憶デバイスから読み出したデータを前記情報処理装置に送信する手段とを備えることを特徴とする記憶デバイス制御装置。

【請求項 5】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有するチャンネル制御部を備える記憶デバイス制御装置であって、

前記情報処理装置から、あるファイルについての前記記憶デバイスの記憶領域上の記憶位置を特定する情報のリクエストを受信して、前記情報を前記情報処理装置に送信する手段と、

前記情報処理装置から前記情報に基づいて生成されたブロック単位でのデータ書き込み要求及び書き込みデータを受信して、これに対応する I / O 要求及び前記書き込みデータを前記記憶デバイスに対して出力する手段とを備えることを特徴とする記憶デバイス制御装置。

【請求項 6】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有する複数のチャンネル制御部を備える記憶デバイス制御装置であって、

前記記憶デバイスの記憶領域上に論理的に設定された論理ボリュームのうち 1 以上の論理ボリュームを、前記各チャンネル制御部からアクセス可能な共有の論理ボリュームに設定する手段と、

前記共有の論理ボリュームに記憶された、ある前記チャンネル制御部の処理を他

の前記チャネル制御部にフェイルオーバさせるための前記各チャネル制御部の引き継ぎ情報に基づいて前記フェイルオーバを実行する手段とを備えることを特徴とする記憶デバイス制御装置。

【請求項 7】 前記フェイルオーバの実行は、

前記情報処理装置から前記フェイルオーバの実行要求を受信した場合、又は前記あるチャネル制御部に障害が発生した場合に行われることを特徴とする請求項 6 に記載の記憶デバイス制御装置。

【請求項 8】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有するチャネル制御部を備える記憶デバイス制御装置であって、

前記ファイルアクセス処理部はアクセス可能な前記情報処理装置の識別情報を記憶し、前記データ入出力要求が、前記識別情報が記憶されている前記情報処理装置から送信されたものである場合にのみそのデータ入出力要求を受け付けることを特徴とする記憶デバイス制御装置。

【請求項 9】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有するチャネル制御部を備える記憶デバイス制御装置の制御方法であって、

前記チャネル制御部が、あるファイルに対する前記データ入出力要求を前記情報処理装置から受信するステップと、

前記ファイルに対する排他制御を実行するステップとを備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項 1 0】 前記データ入出力要求は少なくとも 2 種以上のネットワークファイルシステムプロトコルに従って送信され、あるネットワークファイルシステムプロトコルに従って送信されてくる前記データ入出力要求を受け付けて前記排他制御を実行中に、他のネットワークファイルシステムプロトコルに従って

送信されてくる前記データ入出力要求を受け付けた場合に、当該データ入出力要求についても前記排他制御の効果を反映させることを特徴とする請求項 9 に記載の記憶デバイス制御装置の制御方法。

【請求項 1 1】 前記記憶デバイスの記憶領域は、この記憶領域上に論理的に設定された論理ボリュームを単位として管理され、

前記 I / O プロセッサは、前記ファイルに対する排他制御に対応して前記論理ボリュームに対する排他制御を実行することを特徴とする請求項 9 に記載の記憶デバイス制御装置の制御方法。

【請求項 1 2】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有するチャンネル制御部と、

前記情報処理装置から送信されるブロック単位でのデータ読み出し要求を受信して、これに対応する I / O 要求を前記記憶デバイスに対して出力する手段とを備える記憶デバイス制御装置の制御方法であって、

前記情報処理装置から、あるファイルについての前記記憶デバイスの記憶領域上の記憶位置を特定する情報のリクエストを受信して、前記情報を前記情報処理装置に送信するステップと、

前記情報処理装置から前記情報に基づいて生成されたブロック単位でのデータ読み出し要求を受信するステップと、

前記データ読み出し要求に対応する I / O 要求を前記記憶デバイスに対して出力するステップと、

前記記憶デバイスから読み出したデータを前記情報処理装置に送信するステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項 1 3】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有するチャンネル制御部と、

前記情報処理装置から送信されるブロック単位でのデータ書き込み要求を受信して、これに対応する I / O 要求を前記記憶デバイスに対して出力する手段とを備える記憶デバイス制御装置の制御方法であって、

前記情報処理装置から、あるファイルについての前記記憶デバイスの記憶領域上の記憶位置を特定する情報のリクエストを受信して、前記情報を前記情報処理装置に送信するステップと、

前記情報処理装置から前記情報に基づいて生成されたブロック単位でのデータ書き込み要求及び書き込みデータを受信するステップと、

前記データ書き込み要求に対応する I / O 要求及び前記書き込みデータを前記記憶デバイスに対して出力するステップとを備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項 1 4】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有する複数のチャネル制御部を備える記憶デバイス制御装置の制御方法であって、

前記記憶デバイスの記憶領域上に論理的に設定された論理ボリュームのうち 1 以上の論理ボリュームを、前記各チャネル制御部からアクセス可能な共有の論理ボリュームに設定するステップと、

前記共有の論理ボリュームに記憶された、ある前記チャネル制御部の処理を他の前記チャネル制御部にフェイルオーバーさせるための前記各チャネル制御部の引き継ぎ情報に基づいて前記フェイルオーバーを実行するステップとを備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項 1 5】 前記フェイルオーバーの実行は、

前記情報処理装置から前記フェイルオーバーの実行要求を受信した場合、又は前記あるチャネル制御部に障害が発生した場合に行われることを特徴とする請求項 1 4 に記載の記憶デバイス制御装置の制御方法。

【請求項 1 6】 情報処理装置から送信されるファイル単位でのデータ入出

力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I/O 要求を出力する I/O プロセッサとが形成された回路基板を有するチャネル制御部を備える記憶デバイス制御装置の制御方法であって、

前記ファイルアクセス処理部が、アクセス可能な前記情報処理装置の識別情報を記憶するステップと、

前記データ入出力要求が、前記識別情報が記憶されている前記情報処理装置から送信されたものである場合にのみそのデータ入出力要求を受け付けるステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、記憶デバイス制御装置、及びデバイス制御装置の制御方法に関する。

【0002】

【従来の技術】

近年コンピュータシステムで取り扱われるデータ量が急激に増加している。このようなデータを管理するためのストレージシステムとして、最近ではミッドレンジクラスやエンタープライズクラスと呼ばれるような、巨大な記憶資源を提供する RAID (Redundant Arrays of Inexpensive Disks) 方式で管理された大規模なストレージシステムが注目されている。また、かかる膨大なデータを効率よく利用し管理するために、ディスクアレイ装置等のストレージシステムと情報処理装置とを専用のネットワーク (Storage Area Network、以下 SAN と記す) で接続し、ストレージシステムへの高速かつ大量なアクセスを実現する技術が開発されている。

一方、ストレージシステムと情報処理装置とを TCP/IP (Transmission Control Protocol/Internet Protocol) プロトコル等を用いたネットワークで相互に接続し、情報処理装置からのファイルレベルでのアクセスを実現する NAS

(Network Attached Storage) と呼ばれるストレージシステムが開発されている。

【0003】

【特許文献1】

特開2002-351703号公報

【0004】

【発明が解決しようとする課題】

しかしながら従来のNASは、TCP/IP通信機能及びファイルシステム機能を持たないストレージシステムに、TCP/IP通信機能及びファイルシステム機能を持った情報処理装置を接続させることにより実現されていた。そのため、上記接続される情報処理装置の設置スペースが必要であった。また上記情報処理装置とストレージシステムとの間は、高速に通信を行う必要性からSANで接続されていることが多く、そのための通信制御機器や通信制御機能を備える必要もあった。

本発明は上記課題を鑑みてなされたものであり、記憶デバイス制御装置、及びデバイス制御装置の制御方法を提供することを主たる目的とする。

【0005】

【課題を解決するための手段】

上記課題を解決するために、本発明に係る記憶デバイス制御装置は、情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有するチャネル制御部を備える記憶デバイス制御装置であって、前記チャネル制御部が、あるファイルに対する前記データ入出力要求を前記情報処理装置から受信した場合に、そのファイルに対する排他制御を実行する排他制御手段を備える。

【0006】

なお、前記情報処理装置とは、前記構成の前記記憶デバイス制御装置を備えて構成されるストレージシステムにLANやSANを介してアクセスする、例えば

、パーソナルコンピュータやメインフレームコンピュータである。ファイルアクセス処理部の機能はCPU上で実行されるオペレーティングシステムおよびこのオペレーティングシステム上で動作する例えばNFS (Network File System) 等のソフトウェアによって提供される。記憶デバイスは例えばハードディスク装置などのディスクドライブである。I/Oプロセッサは例えばファイルアクセス処理部のハードウェア要素である前記CPUとは独立したIC (Integrated Circuit) をハードウェア要素とし、ファイルアクセス処理部とディスク制御部との間の通信を制御する。ディスク制御部は、記憶デバイスに対してデータの書き込みや読み出しを行う。

その他、本願が開示する課題、及びその解決方法は、発明の実施の形態の欄、及び図面により明らかにされる。

【0007】

【発明の実施の形態】

以下、本発明の実施の形態について図面を用いて詳細に説明する。

まず、本実施の形態に係るストレージシステム600の全体構成を示すブロック図を図1に示す。

【0008】

===全体構成例===

ストレージシステム600は、記憶デバイス制御装置100と記憶デバイス300とを備えている。記憶デバイス制御装置100は、情報処理装置200から受信したコマンドに従って記憶デバイス300に対する制御を行う。例えば情報処理装置200からデータの入出力要求を受信して、記憶デバイス300に記憶されているデータの入出力のための処理を行う。データは、記憶デバイス300が備えるディスクドライブにより提供される物理的な記憶領域上に論理的に設定される記憶領域である論理ボリューム (Logical Unit) (以下、LUと記す) に記憶されている。また記憶デバイス制御装置100は、情報処理装置200との間で、ストレージシステム600を管理するための各種コマンドの授受も行う。情報処理装置200はCPU (Central Processing Unit) やメモリを備えたコンピュータである。情報処理装置200が備えるCPUにより各種プログラムが

実行されることにより様々な機能が実現される。情報処理装置200は、例えばパーソナルコンピュータやワークステーションであることもあるし、メインフレームコンピュータであることもある。

【0009】

図1において、情報処理装置1乃至3(200)は、LAN(Local Area Network)400を介して記憶デバイス制御装置100と接続されている。LAN400は、インターネットとすることもできるし、専用のネットワークとすることもできる。LAN400を介して行われる情報処理装置1乃至3(200)と記憶デバイス制御装置100との間の通信は、例えばTCP/IPプロトコルに従って行われる。情報処理装置1乃至3(200)からは、ストレージシステム600に対して、ファイル名指定によるデータアクセス要求(ファイル単位でのデータ入出力要求。以下、ファイルアクセス要求と記す)が送信される。

【0010】

LAN400にはバックアップデバイス910が接続されている。バックアップデバイス910は具体的にはMOやCD-R、DVD-RAMなどのディスク系デバイス、DATテープ、カセットテープ、オープンテープ、カートリッジテープなどのテープ系デバイスである。バックアップデバイス910は、LAN400を介して記憶デバイス制御装置100との間で通信を行うことにより、記憶デバイス300に記憶されているデータのバックアップデータを記憶する。またバックアップデバイス910は情報処理装置1(200)と接続されるようにすることもできる。この場合は情報処理装置1(200)を介して記憶デバイス300に記憶されているデータのバックアップデータを取得するようにする。

【0011】

記憶デバイス制御装置100は、チャンネル制御部1乃至4(110)を備える。記憶デバイス制御装置100は、チャンネル制御部1乃至4(110)によりLAN400を介して情報処理装置1乃至3(200)やバックアップデバイス910との間で通信を行う。チャンネル制御部1乃至4(110)は、情報処理装置1乃至3(200)からのファイルアクセス要求を個々に受け付ける。すなわち、チャンネル制御部1乃至4(110)には、個々にLAN400上のネットワー

クアドレス（例えば、IPアドレス）が割り当てられていてそれぞれが個別にNASとして振る舞い、個々のNASがあたかも独立したNASが存在するかのようになりNASとしてのサービスを情報処理装置1乃至3（200）に提供することができる。以下、チャンネル制御部1乃至4（110）をCHNと記す。このように1台のストレージシステム600に個別にNASとしてのサービスを提供するチャンネル制御部1乃至4（110）を備えるように構成したことで、従来、独立したコンピュータで個々に運用されていたNASサーバが一台のストレージシステム600に集約される。そして、これによりストレージシステム600の統括的な管理が可能となり、各種設定・制御や障害管理、バージョン管理といった保守業務の効率化が図られる。

【0012】

なお、本実施の形態に係る記憶デバイス制御装置100のチャンネル制御部1乃至4（110）は、後述するように、一体的にユニット化された回路基板上に形成されたハードウェア及びこのハードウェアにより実行されるオペレーティングシステム（以下、OSと記す）やこのOS上で動作するアプリケーションプログラムなどのソフトウェアにより実現される。このように本実施例のストレージシステム600では、従来ハードウェアの一部として実装されてきた機能がソフトウェアにより実現されている。このため、本実施例のストレージシステム600では柔軟性に富んだシステム運用が可能となり、多様で変化の激しいユーザニーズによりきめ細かなサービスを提供することが可能となる。

【0013】

情報処理装置3乃至4（200）はSAN（Storage Area Network）500を介して記憶デバイス制御装置100と接続されている。SAN500は、記憶デバイス300が提供する記憶領域におけるデータの管理単位であるブロックを単位として情報処理装置3乃至4（200）との間でデータの授受を行うためのネットワークである。SAN500を介して行われる情報処理装置3乃至4（200）と記憶デバイス制御装置100との間の通信は、一般にファイバチャネルプロトコルに従って行われる。情報処理装置3乃至4（200）からは、ストレージシステム600に対して、ファイバチャネルプロトコルに従ってブロック単位

のデータアクセス要求（以下、ブロックアクセス要求と記す）が送信される。

SAN500にはSAN対応のバックアップデバイス900が接続されている。SAN対応バックアップデバイス900は、SAN500を介して記憶デバイス制御装置100との間で通信を行うことにより、記憶デバイス300に記憶されているデータのバックアップデータを記憶する。

【0014】

記憶デバイス制御装置100は、チャンネル制御部5乃至6（110）を備える。記憶デバイス制御装置100は、チャンネル制御部5乃至6（110）によりSAN500を介して情報処理装置3乃至4（200）及びSAN対応バックアップデバイス900との間の通信を行う。以下、チャンネル制御部5乃至6（110）をCHFと記す。

また情報処理装置5（200）は、LAN400やSAN500等のネットワークを介さずに記憶デバイス制御装置100と接続されている。情報処理装置5（200）としては例えばメインフレームコンピュータとすることができる。情報処理装置5（200）と記憶デバイス制御装置100との間の通信は、例えばFICON（Fibre Connection）（登録商標）やESCON（Enterprise System Connection）（登録商標）、ACONARC（Advanced Connection Architecture）（登録商標）、FIBARC（Fibre Connection Architecture）（登録商標）などの通信プロトコルに従って行われる。情報処理装置5（200）からは、ストレージシステム600に対して、これらの通信プロトコルに従ってブロックアクセス要求が送信される。

記憶デバイス制御装置100は、チャンネル制御部7乃至8（110）により情報処理装置5（200）との間で通信を行う。以下、チャンネル制御部7乃至8（110）をCHAと記す。

【0015】

SAN500にはストレージシステム600の設置場所（プライマリサイト）とは遠隔した場所（セカンダリサイト）に設置される他のストレージシステム610が接続している。ストレージシステム610は、後述するレプリケーション又はリモートコピーの機能におけるデータの複製先の装置として利用される。な

お、ストレージシステム610はSAN500以外にもATMなどの通信回線によりストレージシステム600に接続していることもある。この場合には例えばチャンネル制御部110として上記通信回線を利用するためのインタフェース（チャンネルエクステンダ）を備えるチャンネル制御部110が採用される。

【0016】

本実施例によれば、ストレージシステム600内にCHN110、CHF110、CHA110を混在させて装着させることにより、異種ネットワークに接続されるストレージシステムを実現できる。具体的には、ストレージシステム600は、CHN110を用いてLAN140に接続し、かつCHF110を用いてSAN500に接続するという、SAN-NAS統合ストレージシステムである。

【0017】

===記憶デバイス===

記憶デバイス300は多数のディスクドライブ（物理ディスク）を備えており、情報処理装置200に対して記憶領域を提供する。データは、ディスクドライブにより提供される物理的な記憶領域上に論理的に設定される記憶領域であるLUに記憶されている。ディスクドライブとしては、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等様々なものを用いることができる。なお、記憶デバイス300は例えば複数のディスクドライブによりディスクアレイを構成するようにすることもできる。この場合、情報処理装置200に対して提供される記憶領域は、RAIDにより管理された複数のディスクドライブにより提供されるようにすることもできる。記憶デバイス制御装置100と記憶デバイス300との間は図1のように直接に接続される形態とすることもできるし、ネットワークを介して接続するようにすることもできる。さらに記憶デバイス300は記憶デバイス制御装置100と一体として構成されることもできる。

【0018】

記憶デバイス300に設定されるLUには、情報処理装置200からアクセス可能なユーザLUや、チャンネル制御部110の制御のために使用されるシステムLU等がある。システムLUにはCHN110で実行されるオペレーティングシ

システムも格納される。また各LUにはチャンネル制御部110が対応付けられている。これによりチャンネル制御部110毎にアクセス可能なLUが割り当てられている。また上記対応付けは、複数のチャンネル制御部110で一つのLUを共有するようにすることもできる。なお以下において、ユーザLUやシステムLUをユーザディスク、システムディスク等とも記す。また、複数のチャンネル制御部110で共有されるLUを共有LUあるいは共有ディスクと記す。

【0019】

===記憶デバイス制御装置===

記憶デバイス制御装置100はチャンネル制御部110、共有メモリ120、キャッシュメモリ130、ディスク制御部140、管理端末160、接続部150を備える。

【0020】

チャンネル制御部110は情報処理装置200との間で通信を行うための通信インタフェースを備え、情報処理装置200との間でデータ入出力コマンド等を授受する機能を備える。例えばCHN110は情報処理装置1乃至3(200)からのファイルアクセス要求を受け付ける。これによりストレージシステム600はNASとしてのサービスを情報処理装置1乃至3(200)に提供することができる。またCHF110は情報処理装置3乃至4(200)からのファイバチャネルプロトコルに従ったブロックアクセス要求を受け付ける。これによりストレージシステム600は高速アクセス可能なデータ記憶サービスを情報処理装置3乃至4(200)に対して提供することができる。またCHA110は情報処理装置5(200)からのFICONやESCON、ACONARC、FIBARC等のプロトコルに従ったブロックアクセス要求を受け付ける。これによりストレージシステム600は情報処理装置5(200)のようなメインフレームコンピュータに対してもデータ記憶サービスを提供することができる。

【0021】

各チャンネル制御部110は管理端末160と共に内部LAN151で接続されている。これによりチャンネル制御部110に実行させるマイクロプログラム等を管理端末160から送信しインストールすることが可能となっている。チャンネル

制御部 110 の構成については後述する。

【0022】

接続部 150 はチャネル制御部 110、共有メモリ 120、キャッシュメモリ 130、ディスク制御部 140 を相互に接続する。チャネル制御部 110、共有メモリ 120、キャッシュメモリ 130、ディスク制御部 140 間でのデータやコマンドの授受は接続部 150 を介することにより行われる。接続部 150 は例えば高速スイッチングによりデータ伝送を行う超高速クロスバススイッチなどの高速バスで構成される。チャネル制御部 110 同士が高速バスで接続されていることで、個々のコンピュータ上で動作する NAS サーバが LAN を通じて接続する従来の構成に比べてチャネル制御部 110 間の通信パフォーマンスが大幅に向上している。またこれにより高速なファイル共有機能や高速フェイルオーバーなどが可能となる。

【0023】

共有メモリ 120 及びキャッシュメモリ 130 は、チャネル制御部 110、ディスク制御部 140 により共有される記憶メモリである。共有メモリ 120 は主に制御情報やコマンド等を記憶するために利用されるのに対し、キャッシュメモリ 130 は主にデータを記憶するために利用される。

例えば、あるチャネル制御部 110 が情報処理装置 200 から受信したデータ入出力コマンドが書き込みコマンドであった場合には、当該チャネル制御部 110 は書き込みコマンドを共有メモリ 120 に書き込むと共に、情報処理装置 200 から受信した書き込みデータをキャッシュメモリ 130 に書き込む。一方、ディスク制御部 140 は共有メモリ 120 を監視しており、共有メモリ 120 に書き込みコマンドが書き込まれたことを検出すると、当該コマンドに従ってキャッシュメモリ 130 から書き込みデータを読み出して記憶デバイス 300 に書き込む。

【0024】

またあるチャネル制御部 110 が情報処理装置 200 から受信したデータ入出力コマンドが読み出しコマンドであった場合には、当該チャネル制御部 110 は読み出しコマンドを共有メモリ 120 に書き込むと共に、読み出し対象となるデ

ータがキャッシュメモリ 1 3 0 に存在するかどうかを調べる。ここでキャッシュメモリ 1 3 0 に存在すれば、チャンネル制御部 1 1 0 はそのデータを情報処理装置 2 0 0 に送信する。一方、読みだし対象となるデータがキャッシュメモリ 1 3 0 に存在しない場合には、共有メモリ 1 2 0 を監視することにより読み出しコマンドが共有メモリ 1 2 0 に書き込まれたことを検出したディスク制御部 1 4 0 が、記憶デバイス 3 0 0 から読みだし対象となるデータを読み出してこれをキャッシュメモリ 1 3 0 に書き込むと共に、その旨を共有メモリ 1 2 0 に書き込む。そして、チャンネル制御部 1 1 0 は共有メモリ 1 2 0 を監視することにより読みだし対象となるデータがキャッシュメモリ 1 3 0 に書き込まれたことを検出すると、そのデータを情報処理装置 2 0 0 に送信する。

【 0 0 2 5 】

なお、このようにチャンネル制御部 1 1 0 からディスク制御部 1 4 0 に対するデータの書き込みや読み出しの指示を共有メモリ 1 2 0 を介在させて間接に行う構成の他、例えばチャンネル制御部 1 1 0 からディスク制御部 1 4 0 に対してデータの書き込みや読み出しの指示を共有メモリ 1 2 0 を介さずに直接に行う構成とすることもできる。

【 0 0 2 6 】

ディスク制御部 1 4 0 は記憶デバイス 3 0 0 の制御を行う。例えば上述のように、チャンネル制御部 1 1 0 が情報処理装置 2 0 0 から受信したデータ書き込みコマンドに従って記憶デバイス 3 0 0 へデータの書き込みを行う。また、チャンネル制御部 1 1 0 により送信された論理アドレス指定による LU へのデータアクセス要求を、物理アドレス指定による物理ディスクへのデータアクセス要求に変換する。記憶デバイス 3 0 0 における物理ディスクが R A I D により管理されている場合には、R A I D 構成に従ったデータのアクセスを行う。またディスク制御部 1 4 0 は、記憶デバイス 3 0 0 に記憶されたデータの複製管理の制御やバックアップ制御を行う。さらにディスク制御部 1 4 0 は、災害発生時のデータ消失防止（ディザスタリカバリ）などを目的としてプライマリサイトのストレージシステム 6 0 0 のデータの複製をセカンダリサイトに設置された他のストレージシステム 6 1 0 にも記憶する制御（レプリケーション機能、又はリモートコピー機能）

などを行う。

【0027】

各ディスク制御部140は管理端末160と共に内部LAN151で接続されており、相互に通信を行うことが可能である。これにより、ディスク制御部140に実行させるマイクロプログラム等を管理端末160から送信しインストールすることが可能となっている。ディスク制御部140の構成については後述する。

【0028】

本実施例においては、共有メモリ120及びキャッシュメモリ130がチャンネル制御部110及びディスク制御部140に対して独立に設けられていることについて記載したが、本実施例はこの場合に限られるものでなく、共有メモリ120又はキャッシュメモリ130がチャンネル制御部110及びディスク制御部140の各々に分散されて設けられることも好ましい。この場合、接続部150は、分散された共有メモリ又はキャッシュメモリを有するチャンネル制御部110及びディスク制御部140を相互に接続させることになる。

【0029】

===管理端末===

管理端末160はストレージシステム600を保守・管理するためのコンピュータである。管理端末160を操作することにより、例えば記憶デバイス300内の物理ディスク構成の設定や、LUの設定、チャンネル制御部110において実行されるマイクロプログラムのインストール等を行うことができる。ここで、記憶デバイス300内の物理ディスク構成の設定としては、例えば物理ディスクの増設や減設、RAID構成の変更（RAID1からRAID5への変更等）等を行うことができる。さらに管理端末160からは、ストレージシステム600の動作状態の確認や故障部位の特定、チャンネル制御部110で実行されるオペレーティングシステムのインストール等の作業を行うこともできる。また管理端末160はLANや電話回線等で外部保守センタと接続されており、管理端末160を利用してストレージシステム600の障害監視を行ったり、障害が発生した場合に迅速に対応することも可能である。障害の発生は例えばOSやアプリケーション

ョンプログラム、ドライバソフトウェアなどから通知される。この通知はHTTPプロトコルやSNMP (Simple Network Management Protocol)、電子メールなどにより行われる。これらの設定や制御は、管理端末160で動作するWebサーバが提供するWebページをユーザインタフェースとしてオペレータなどにより行われる。オペレータ等は、管理端末160を操作して障害監視する対象や内容の設定、障害通知先の設定などを行うこともできる。

【0030】

管理端末160は記憶デバイス制御装置100に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。また管理端末160は、記憶デバイス制御装置100及び記憶デバイス300の保守・管理を専用に行うコンピュータとすることもできるし、汎用のコンピュータに保守・管理機能を持たせたものとすることもできる。

【0031】

管理端末160の構成を示すブロック図を図2に示す。

管理端末160は、CPU161、メモリ162、ポート163、記録媒体読取装置164、入力装置165、出力装置166、記憶装置168を備える。

CPU161は管理端末160の全体の制御を司るもので、メモリ162に格納されたプログラム162cを実行することにより上記Webサーバとしての機能等を実現する。メモリ162には、物理ディスク管理テーブル162aとLU管理テーブル162bとプログラム162cとが記憶されている。

【0032】

物理ディスク管理テーブル162aは、記憶デバイス300に備えられる物理ディスク（ディスクドライブ）を管理するためのテーブルである。物理ディスク管理テーブル162aを図3に示す。図3においては、記憶デバイス300が備える多数の物理ディスクのうち、ディスク番号#001乃至#006までが示されている。それぞれの物理ディスクに対して、容量、RAID構成、使用状況が示されている。

LU管理テーブル162bは、上記物理ディスク上に論理的に設定されるLUを管理するためのテーブルである。LU管理テーブル162bを図4に示す。図

4 においては、記憶デバイス 3 0 0 上に設定される多数の LU のうち、LU 番号 # 1 乃至 # 3 までが示されている。それぞれの LU に対して、物理ディスク番号、容量、RAID 構成が示されている。

【 0 0 3 3 】

記録媒体読取装置 1 6 4 は、記録媒体 1 6 7 に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ 1 6 2 や記憶装置 1 6 8 に格納される。従って、例えば記録媒体 1 6 7 に記録されたプログラム 1 6 2 c を、記録媒体読取装置 1 6 4 を用いて上記記録媒体 1 6 7 から読み取って、メモリ 1 6 2 や記憶装置 1 6 8 に格納するようにすることができる。記録媒体 1 6 7 としてはフレキシブルディスクや CD-ROM、半導体メモリ等を用いることができる。記録媒体読取装置 1 6 4 は管理端末 1 6 0 に内蔵されている形態とすることもできるし、外付されている形態とすることもできる。記憶装置 1 6 8 は、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等である。入力装置 1 6 5 はオペレータ等による管理端末 1 6 0 へのデータ入力等のために用いられる。入力装置 1 6 5 としては例えばキーボードやマウス等が用いられる。出力装置 1 6 6 は情報を外部に出力するための装置である。出力装置 1 6 6 としては例えばディスプレイやプリンタ等が用いられる。ポート 1 6 3 は内部 LAN 1 5 1 に接続されており、これにより管理端末 1 6 0 はチャネル制御部 1 1 0 やディスク制御部 1 4 0 等と通信を行うことができる。またポート 1 6 3 は、LAN 4 0 0 に接続するようにすることもできるし、電話回線に接続するようにすることもできる。

【 0 0 3 4 】

=== 外観図 ===

次に、本実施の形態に係るストレージシステム 6 0 0 の外観構成を図 5 に示す。また、記憶デバイス制御装置 1 0 0 の外観構成を図 6 に示す。

図 5 に示すように、本実施の形態に係るストレージシステム 6 0 0 は記憶デバイス制御装置 1 0 0 及び記憶デバイス 3 0 0 がそれぞれの筐体に納められた形態をしている。記憶デバイス制御装置 1 0 0 の筐体の両側に記憶デバイス 3 0 0 の筐体が配置されている。

記憶デバイス制御装置100は、正面中央部に管理端末160が備えられている。管理端末160はカバーで覆われており、図6に示すようにカバーを開けることにより管理端末160を使用することができる。なお図6に示した管理端末160はいわゆるノート型パーソナルコンピュータの形態をしているが、どのような形態とすることも可能である。

【0035】

管理端末160の下部には、チャンネル制御部110を装着するためのスロットが設けられている。各スロットにはチャンネル制御部110のボードが装着される。本実施の形態に係るストレージシステム600においては、スロットは8つあり、図5及び図6には8つのスロットにチャンネル制御部110が装着された状態が示されている。各スロットにはチャンネル制御部110を装着するためのガイドレールが設けられている。ガイドレールに沿ってチャンネル制御部110をスロットに挿入することにより、チャンネル制御部110を記憶デバイス制御装置100に装着することができる。また各スロットに装着されたチャンネル制御部110は、ガイドレールに沿って手前方向に引き抜くことにより取り外すことができる。また各スロットの奥手方向正面部には、各チャンネル制御部110を記憶デバイス制御装置100と電氣的に接続するためのコネクタが設けられている。チャンネル制御部110には、CHN、CHF、CHAがあるが、いずれのチャンネル制御部110もサイズやコネクタの位置、コネクタのピン配列等に互換性をもたせているため、8つのスロットにはいずれのチャンネル制御部110も装着することが可能である。従って、例えば8つのスロット全てにCHN110を装着するようにすることもできる。また図1に示したように、4枚のCHN110と、2枚のCHF110と、2枚のCHA110とを装着するようにすることもできる。チャンネル制御部110を装着しないスロットを設けるようにすることもできる。

【0036】

各スロットに装着されたチャンネル制御部110は、同種の複数のチャンネル制御部110でクラスタを構成する。例えば2枚のCHN110をペアとしてクラスタを構成することができる。クラスタを構成することにより、クラスタ内のあるチャンネル制御部110に障害が発生した場合でも、障害が発生したチャンネル制御

部110がそれまで行っていた処理をクラスタ内の他のチャネル制御部110に引き継ぐようにすることができる（フェイルオーバー制御）。2枚のCHN110でクラスタを構成している様子を示す図を図12に示すが、詳細は後述する。

【0037】

なお、記憶デバイス制御装置100は信頼性向上のため電源供給が2系統化されており、チャネル制御部110が装着される上記8つのスロットは電源系統毎に4つずつに分けられている。そこでクラスタを構成する場合には、両方の電源系統のチャネル制御部110を含むようにする。これにより、片方の電源系統に障害が発生し電力の供給が停止しても、同一クラスタを構成する他方の電源系統に属するチャネル制御部110への電源供給は継続されるため、当該チャネル制御部110に処理を引き継ぐ（フェイルオーバー）ことができる。

【0038】

なお、上述したように、チャネル制御部110は上記各スロットに装着可能なボード、すなわち同一のユニットに形成された一つのユニットとして提供されるが、上記同一のユニットは複数枚数の基板から構成されているようにすることもできる。つまり、複数枚数の基板から構成されていても、各基板が相互に接続されて一つのユニットとして構成され、記憶デバイス制御装置100のスロットに対して一体的に装着できる場合は、同一の回路基板の概念に含まれる。

【0039】

ディスク制御部140や共有メモリ120等の、記憶デバイス制御装置100を構成する他の装置については図5及び図6には示されていないが、記憶デバイス制御装置100の背面側等に装着されている。

また記憶デバイス制御装置100には、チャネル制御部110等から発生する熱を放出するためのファン170が設けられている。ファン170は記憶デバイス制御装置100の上面部に設けられる他、チャネル制御部110用スロットの上部にも設けられている。

ところで、筐体に収容されて構成される記憶デバイス制御装置100および記憶デバイス300としては、例えばSAN対応として製品化されている従来構成の装置を利用することができる。特に上記のようにCHNのコネクタ形状を従来

構成の筐体に設けられているスロットにそのまま装着できる形状とすることで従来構成の装置をより簡単に利用することができる。つまり本実施例のストレージシステム600は、既存の製品を利用することで容易に構築することができる。

【0040】

===チャネル制御部===

本実施の形態に係るストレージシステム600は、上述の通りCHN110により情報処理装置1乃至3(200)からのファイルアクセス要求を受け付け、NASとしてのサービスを情報処理装置1乃至3(200)に提供する。

CHN110のハードウェア構成を図7に示す。この図に示すようにCHN110のハードウェアは一つのユニットで構成される。以下、このユニットのことをNASボードと記す。NASボードは一枚もしくは複数枚の回路基板を含んで構成される。より具体的には、NASボードは、ネットワークインタフェース部111、CPU112、メモリ113、入出力制御部114、I/O(Input/Output)プロセッサ119、NVRAM(Non Volatile RAM)115、ボード接続用コネクタ116、通信コネクタ117を備え、これらが同一のユニットに形成されて構成されている。

【0041】

ネットワークインタフェース部111は、情報処理装置200との間で通信を行うための通信インタフェースを備えている。CHN110の場合は、例えばTCP/IPプロトコルに従って情報処理装置200から送信されたファイルアクセス要求を受信する。通信コネクタ117は情報処理装置200と通信を行うためのコネクタである。CHN110の場合はLAN400に接続可能なコネクタであり、例えばイーサネット(登録商標)に対応している。

CPU112は、CHN110をNASボードとして機能させるための制御を司る。

【0042】

メモリ113には様々なプログラムやデータが記憶される。例えば図9に示すメタデータ730やロックテーブル720、また図11に示すNASマネージャ706等の各種プログラムが記憶される。メタデータ730はファイルシステム

が管理しているファイルに対応させて生成される情報である。メタデータ730には例えばファイルのデータが記憶されているLU上のアドレスやデータサイズなど、ファイルの保管場所を特定するための情報が含まれる。メタデータ730にはファイルの容量、所有者、更新時刻等の情報が含まれることもある。また、メタデータ730はファイルだけでなくディレクトリに対応させて生成されることもある。メタデータ730の例を図13に示す。メタデータ730は記憶デバイス300上の各LUにも記憶されている。

【0043】

ロックテーブル720は、情報処理装置1乃至3(200)からのファイルアクセスに対して排他制御を行うためのテーブルである。排他制御を行うことにより情報処理装置1乃至3(200)でファイルを共用することができる。ロックテーブル720を図14に示す。図14に示すようにロックテーブル720にはファイルロックテーブル721とLUロックテーブル722とがある。ファイルロックテーブル721は、ファイル毎にロックが掛けられているか否かを示すためのテーブルである。いずれかの情報処理装置200によりあるファイルがオープンされている場合に当該ファイルにロックが掛けられる。ロックが掛けられたファイルに対する他の情報処理装置200によるアクセスは禁止される。LUロックテーブル722は、LU毎にロックが掛けられているか否かを示すためのテーブルである。いずれかの情報処理装置200により、あるLUに対するアクセスが行われている場合に当該LUにロックが掛けられる。ロックが掛けられたLUに対する他の情報処理装置200によるアクセスは禁止される。

【0044】

入出力制御部114は、ディスク制御部140やキャッシュメモリ130、共有メモリ120、管理端末160との間でデータやコマンドの授受を行う。入出力制御部114はI/Oプロセッサ119やNVRAM115を備えている。I/Oプロセッサ119は例えば1チップのマイコンで構成される。I/Oプロセッサ119は上記データやコマンドの授受を制御し、CPU112とディスク制御部140との間の通信を中継する。NVRAM115はI/Oプロセッサ119の制御を司るプログラムを格納する不揮発性メモリである。NVRAM115

に記憶されるプログラムの内容は、管理端末160や、後述するNASマネージャ706からの指示により書き込みや書き換えを行うことができる。

【0045】

次にCHF110及びCHA110の構成を示す図を図8に示す。CHF110及びCHA110も、CHN110と同様に一つのユニットとして形成されている。CHN110と同様、上記ユニットは複数枚数の回路基板から構成されているようにすることもできる。またCHF110及びCHA110は、CHN110とサイズやボード接続用コネクタ116の位置、ボード接続用コネクタ116のピン配列等に互換性をもたせている。

CHF110及びCHA110は、ネットワークインタフェース部111、メモリ113、入出力制御部114、I/Oプロセッサ119、NVRAM115、ボード接続用コネクタ116、通信コネクタ117を備える。

【0046】

ネットワークインタフェース部111は、情報処理装置200との間で通信を行うための通信インタフェースを備えている。CHF110の場合は、例えばファイバチャネルプロトコルに従って情報処理装置200から送信されたブロックアクセス要求を受信する。CHA110の場合は、例えばFICON（登録商標）やESCON（登録商標）、ACONARC（登録商標）、FIBARC（登録商標）のプロトコルに従って情報処理装置200から送信されたブロックアクセス要求を受信する。通信コネクタ117は情報処理装置200と通信を行うためのコネクタである。CHF110の場合はSAN500に接続可能なコネクタであり、例えばファイバチャネルに対応している。CHA110の場合は情報処理装置5と接続可能なコネクタであり、FICON（登録商標）やESCON（登録商標）、ACONARC（登録商標）、FIBARC（登録商標）に対応している。

【0047】

入出力制御部114は、それぞれCHF110、CHA110の全体の制御を司ると共に、ディスク制御部140やキャッシュメモリ130、共有メモリ120、管理端末160との間でデータやコマンドの授受を行う。メモリ113に格

納された各種プログラムを実行することにより本実施の形態に係るCHF110及びCHA110の機能が実現される。入出力制御部114はI/Oプロセッサ119やNVRAM115を備えている。I/Oプロセッサ119は上記データやコマンドの授受を制御する。NVRAM115はI/Oプロセッサ119の制御を司るプログラムを格納する不揮発性メモリである。NVRAM115に記憶されるプログラムの内容は、管理端末160や、後述するNASマネージャ706からの指示により書き込みや書き換えを行うことができる。

【0048】

次にディスク制御部140の構成を示す図を図10に示す。

ディスク制御部140は、インタフェース部141、メモリ143、CPU142、NVRAM144、ボード接続用コネクタ145を備え、これらが一体的なユニットとして形成されている。

インタフェース部141は、接続部150を介してチャンネル制御部110等との間で通信を行うための通信インタフェースや、記憶デバイス300との間で通信を行うための通信インタフェースを備えている。

CPU142は、ディスク制御部140全体の制御を司ると共に、チャンネル制御部140や記憶デバイス300、管理端末160との間の通信を行う。メモリ143やNVRAM144に格納された各種プログラムを実行することにより本実施の形態に係るディスク制御部140の機能が実現される。ディスク制御部140により実現される機能としては、記憶デバイス300の制御やRAID制御、記憶デバイス300に記憶されたデータの複製管理やバックアップ制御、リモートコピー制御等である。

NVRAM144はCPU142の制御を司るプログラムを格納する不揮発性メモリである。NVRAM144に記憶されるプログラムの内容は、管理端末160や、NASマネージャ706からの指示により書き込みや書き換えを行うことができる。

またディスク制御部140はボード接続用コネクタ145を備えている。ボード接続用コネクタ145が記憶デバイス制御装置100側のコネクタと嵌合することにより、ディスク制御部140は記憶デバイス制御装置100と電氣的に接

続される。

【0049】

次に、本実施の形態に係るストレージシステム600におけるソフトウェア構成図を図11に示す。

オペレーティングシステム701上では、RAIDマネージャ708、ボリュームマネージャ707、SVPマネージャ709、ファイルシステムプログラム703、ネットワーク制御部702、バックアップ管理プログラム710、障害管理プログラム705、NASマネージャ706などのソフトウェアが動作する。

【0050】

オペレーティングシステム701上で動作するRAIDマネージャ708は、RAID制御部740に対するパラメータの設定やRAID制御部740を制御する機能を提供する。RAIDマネージャ708はオペレーティングシステム701やオペレーティングシステム701上で動作する他のアプリケーション、もしくはSVPからパラメータや制御指示情報を受け付けて、受け付けたパラメータのRAID制御部740への設定や、RAID制御部指示情報に対応する制御コマンドの送信を行う。

【0051】

ここで設定されるパラメータとしては、例えば、RAIDグループを構成する記憶デバイス（物理ディスク）を定義（RAIDグループの構成情報、ストライプサイズの指定など）するためのパラメータ、RAIDレベル（例えば0, 1, 5）を設定するためのパラメータなどがある。また、RAIDマネージャ708がRAID制御部740に送信する制御コマンドとしてはRAIDの構成・削除・容量変更を指示するコマンド、各RAIDグループの構成情報を要求するコマンドなどがある。

【0052】

ボリュームマネージャ707は、RAID制御部740によって提供されるLUをさらに仮想化した仮想化論理ボリュームをファイルシステムプログラム703に提供する。1つの仮想化論理ボリュームは1以上の論理ボリュームによって

構成される。

【0053】

ファイルシステムプログラム703の主な機能は、ネットワーク制御部702が受信したファイルアクセス要求に指定されているファイル名とそのファイル名が格納されている仮想化論理ボリューム上のアドレスとの対応づけを管理することである。例えば、ファイルシステムプログラム703はファイルアクセス要求に指定されているファイル名に対応する仮想化論理ボリューム上のアドレスを特定する。

【0054】

ネットワーク制御部702は、NFS (Network File System) 711とSamba 712の2つのファイルシステムプロトコルを含んで構成される。NFS 711は、NFS 711が動作するUNIX (登録商標) 系の情報処理装置200からのファイルアクセス要求を受け付ける。一方、Samba 712はCIFS (Common Interface File System) 713が動作するWindows (登録商標) 系の情報処理装置200からのファイルアクセス要求を受け付ける。

【0055】

NASマネージャ706は、ストレージシステム600について、その動作状態の確認、設定や制御などを行うためのプログラムである。NASマネージャ706はWebサーバとしての機能も有し、情報処理装置200からストレージシステム600の設定や制御を行うための設定Webページを情報処理装置200に提供する。NASマネージャ706は、情報処理装置1乃至3 (200) からのHTTP (HyperText Transport Protocol) リクエストに応じて、設定Webページのデータを情報処理装置1乃至3 (200) に送信する。情報処理装置1乃至3 (200) に表示された設定Webページを利用してシステムアドミニストレータなどによりストレージシステム600の設定や制御の指示が行われる。設定Webページを利用して行うことができる内容としては、例えば、LUの管理や設定 (容量管理や容量拡張・縮小、ユーザ割り当て等)、上述の複製管理やリモートコピー (レプリケーション) 等の機能に関する設定や制御 (複製元のLUと複製先のLUの設定など)、後述のバックアップ管理プログラム710につ

いての設定や制御、冗長構成されたCHNやCHF、CHA間でのクラスタの管理（フェイルオーバーさせる相手の対応関係の設定、フェイルオーバー方法など）、OSやOS上で動作するアプリケーションプログラムのバージョン管理、ウイルス検知プログラムやウイルス駆除などのデータの安全性に関する機能を提供するセキュリティ管理プログラム716の動作状態の管理や設定および更新管理（バージョン管理）などがある。NASマネージャ706は、設定Webページに対する操作に起因して情報処理装置200から送信される設定や制御に関するデータを受信してそのデータに対応する設定や制御を実行する。これにより、情報処理装置1乃至3（200）からストレージシステム600の様々な設定や制御を行うことができる。

【0056】

バックアップ管理プログラム710は、記憶デバイス300に記憶されているデータをLAN経由またはSAN経由でバックアップするためのプログラムである。バックアップ管理プログラム710はNDMP（Network Data Management Protocol）の機能を提供し、情報処理装置200で動作するNDMP対応のバックアップソフトウェアとLAN400を通じてNDMPに従った通信を行う。バックアップデバイス910が情報処理装置200にSCSI経由などで接続されている場合、バックアップされるデータは情報処理装置200に一旦取り込まれてからバックアップデバイス910に送られる。バックアップデバイス910がLANに200接続されている場合には、バックアップされるデータを、情報処理装置200を経由せずにストレージシステム600から直接バックアップデバイス910に転送することもできる。

【0057】

障害管理プログラム705は、クラスタを構成するチャネル制御部110間でのフェイルオーバー制御を行うためのプログラムである。

SVPマネージャ709は、管理端末160からの要求に応じて各種のサービスを管理端末160に提供する。例えば、LUの設定内容やRAIDの設定内容等のストレージシステム600に関する各種設定内容の管理端末160への提供や、管理端末160から入力されたストレージシステム600に関する各種設定

の反映等を行う。

セキュリティ管理プログラム716は、コンピュータウイルスの検知、侵入監視、コンピュータウイルス検知プログラムの更新管理、感染したコンピュータウイルスの駆除、ファイアウォール機能などを実現する。

【0058】

次に、2枚のCHN110でクラスタ180が構成されている様子を示す図を図12に示す。図12では、CHN1（チャンネル制御部1）110とCHN2（チャンネル制御部2）110とでクラスタ180が構成されている場合を示す。

上述したように、フェイルオーバー処理はクラスタ180を構成するチャンネル制御部110間で行われる。つまり、例えばCHN1（110）に何らかの障害が発生し処理を継続することができなくなった場合には、CHN1（110）がそれまで行っていた処理はCHN2（110）に引き継がれる。フェイルオーバー処理は、CHN1（110）とCHN2（110）により実行される障害管理プログラム705により実行される。

【0059】

CHN1（110）及びCHN2（110）は共に障害管理プログラム705を実行し、共有メモリ120に対して自己の処理が正常に行われていることを書き込むようにする。そして、相手側の上記書き込みの有無を相互に確認するようにする。相手側による書き込みが検出できない場合には、相手側に何らかの障害が発生したと判断し、フェイルオーバー処理を実行する。フェイルオーバー実行時の処理の引き継ぎは、共有LU310を介して行われる。

【0060】

また、クラスタ180を構成する各CHN110のファイルアクセス処理部は、アクセス可能な情報処理装置1乃至3（200）を管理することができる。これにより、アクセス可能な情報処理装置1乃至3（200）からファイルアクセス要求が送信されてきた場合にのみ、そのファイルアクセス要求を受け付けるようにすることができる。アクセス可能な情報処理装置1乃至3（200）の管理は、各CHN110のメモリ113に予めアクセスを許可する情報処理装置1乃至3（200）のIPアドレスのドメイン名（識別情報）を記憶させることによ

り行う。

【0061】

これにより、情報処理装置1乃至3(200)が共通のLAN400を介してストレージシステム600に接続されていても、情報処理装置1乃至3(200)のそれぞれに対して排他的にLUを割り当てることができる。例えば情報処理装置1乃至3(200)がそれぞれ別の企業のコンピュータである場合に、情報処理装置1乃至3(200)のそれぞれに対して相互にデータの秘匿性を保った記憶サービスを提供することができる。

【0062】

各CHN110に対する上記設定は、管理端末160や情報処理装置1乃至3(200)から行うことができる。情報処理装置1乃至3(200)から行う場合には、CHN110上で動作するNASマネージャ706により情報処理装置1乃至3(200)に表示される設定Webページを利用して行う。

【0063】

===ファイルデータの排他制御===

次に本実施の形態に係るファイルデータの排他制御について説明する。前述したように排他制御にはファイル単位での排他制御とLU単位での排他制御とがある。これらの排他制御を行うことによりファイル更新の順序性を確保することができるので、情報処理装置1乃至3(200)間でファイルを共有させるようにすることができるようになる。本実施の形態に係る排他制御は、各種の動作を行うためのコードから構成されるネットワークファイルシステムプログラムを、CHN110が備えるCPU112やI/Oプロセッサ119が実行することにより実現される。ここでネットワークファイルシステムプログラムはNFS711やSamba712等のファイルシステムプロトコルを制御するためのプログラムである。

【0064】

まず本実施の形態に係るファイル単位での排他制御を説明するためのフローチャートを図15に示す。

CHN110のネットワークインタフェース部111は、情報処理装置1乃至

3 (200) からファイルアクセス要求 (データアクセスコマンド) を受信する (S2000)。ファイルアクセス要求には、ファイル名や、読み出し又は書き込み等のファイルアクセスの種別、書き込みの場合には書き込みデータ、LAN 400 の通信プロトコルのヘッダ情報等が含まれている。その後 CPU 112 はネットワークインタフェース部 111 で受信したファイルアクセス要求からファイル名を抽出する (S2001)。そして CPU 112 はメモリ 113 に記憶されているファイルロックテーブル 721 を参照し、ファイルアクセス要求から抽出されたファイル名に基づいて、当該ファイルのロック状態をチェックする (S2002)。

【0065】

ロック中である場合には (S2003)、他の情報処理装置 200 により当該ファイルにアクセスが行われているので、そのファイルに対するアクセスは行えない。そこで、その旨のメッセージをファイルアクセス要求を送信してきた情報処理装置 200 に対して送信する (S2007)。なお、ファイルアクセス要求のあったファイルがロック中であつた場合に、当該ファイルへのアクセスを一律に禁止するのではなく、アクセスの種別に応じて、例えば書き込みの場合にのみ禁止するようにすることもできる。また CHN 110 は情報処理装置 200 にファイルアクセス禁止のメッセージを送信するだけでなく、その後当該ファイルのロックが解除された際にはその旨のメッセージを情報処理装置 200 に送信するようにすることもできる。ファイルアクセス禁止のメッセージを受信した情報処理装置 200 は、当該ファイルに対するアクセスを取りやめるようにすることもできるし、所定の時間が経過した後に当該ファイルへのアクセスを再度行うようにすることもできる。さらには CHN 110 からのロック解除のメッセージを受けてから再度ファイルアクセスを行うようにすることもできる。

【0066】

情報処理装置 200 からアクセス要求のあったファイルがロック中でなかった場合には (S2003)、CPU 112 はファイルロックテーブル 721 の当該ファイルをロック中にする (S2004)。これにより当該ファイルに対する他の情報処理装置 200 によるアクセスは禁止される。そして CPU 112 はメモリ 113 に記憶されているメタデータ 730 を参照し、当該ファイルのデータの先頭記憶

位置とデータ長（容量）を取得する（S2005）。そして入出力制御部 1 1 4 に対してディスクアクセスを指示する（S2006）。入出力制御部 1 1 4 は当該ファイルの記憶位置とデータ長に基づき、ファイルアクセス要求に対する I/O 要求を生成し、ディスク制御部 1 4 0 に出力する。これによりデータアクセスすなわちデータの読み出しや書き込み等が行われる。データアクセスが完了したら当該ファイルに対するロックを解除する。

なお S2002 においてファイルロックテーブル 7 2 1 をチェックした結果、情報処理装置 2 0 0 からアクセス要求のあったファイルがファイルロックテーブル 7 2 1 に登録されていなかった場合には、当該ファイルをファイルロックテーブル 7 2 1 に追加し S2004 以降の処理を行う。

【 0 0 6 7 】

入出力制御部 1 1 4 の I/O プロセッサ 1 1 9 は、CPU 1 1 2 から受けたディスクアクセス指示に従って I/O 要求をディスク制御部 1 4 0 に出力する。その際、LU 単位での排他制御が行われる。LU 単位での排他制御を説明するためのフローチャートを図 1 6 に示す。

まず I/O プロセッサ 1 1 9 は、CPU 1 1 2 からディスクアクセス指示を受け付ける（S1000）。次に I/O プロセッサ 1 1 9 はメモリ 1 1 3 に記憶されている LU ロックテーブル 7 2 2 を参照し、これからアクセスしようとしている LU 3 1 0 のロック状態をチェックする（S1001）。ロック中である場合にはロックが解除されるまで待つ（S1002）。ロックが解除されたら LU ロックテーブル 7 2 2 の当該 LU をロック中にする（S1003）。これにより当該 LU に対する他のアクセスは禁止される。そして I/O プロセッサ 1 1 9 は I/O 要求を生成し、ディスク制御部 1 4 0 に出力する（S1004）。I/O 要求にはデータの先頭アドレス、データ長、読み出し又は書き込み等のアクセスの種別、書き込みの場合には書き込みデータ等が含まれている。データアクセスが完了したら当該 LU のロックを解除する。

【 0 0 6 8 】

また、ファイルロックテーブル 7 2 1 や LU ロックテーブル 7 2 2 を用いてファイルの排他制御を行うことにより、情報処理装置 1 乃至 3（2 0 0）から送信

されるファイルアクセス要求が、異なるネットワークファイルシステムプロトコルに従っている場合でも、排他制御の効果を各ネットワークファイルシステムプロトコルに従ったファイルアクセス要求に反映させることができる。

【0069】

以上の排他制御がCHN110において行われる。本実施の形態に係るストレージシステム600における記憶デバイス制御装置100にはCHN110、CHF110、CHA110を混在させて装着することが可能であるが、このような構成の下において、本実施の形態に係るストレージシステム600に接続される情報処理装置1乃至3(200)間でのファイルデータの共有を実現することができる。

【0070】

===フェイルオーバー制御===

次に本実施の形態に係るフェイルオーバー制御について説明する。フェイルオーバー制御は、各種の動作を行うためのコードから構成される障害管理プログラム705を、CHN110が備えるCPU112やI/Oプロセッサ119が実行することにより実現される。

図12で示したようにフェイルオーバー制御はクラスタ180を構成するCHN110間で行われる。クラスタ180の設定はNASマネージャ706により行うことができる。フェイルオーバー制御は、CHN110に障害が発生した場合の他、NASマネージャ706からの指示(フェイルオーバーの実行要求)によっても行われる。

【0071】

本実施の形態に係るフェイルオーバー制御を説明するためのフローチャートを図17に示す。

CPU112はNASマネージャ706からフェイルオーバーの指示(実行要求)を受信するとフェイルオーバー処理を開始する(S3000)。

NASマネージャ706からフェイルオーバー指示があったかどうかのチェックは、図17に示す様に障害管理プログラム705の処理ルーチン内で行うようにすることもできるし、NASマネージャ706からフェイルオーバー指示があった

場合に割り込み信号を発生させることにより行うようにすることもできる。

NAS マネージャ 706 からのフェイルオーバ指示がない場合には、CPU 112 は、自己の CHN 110 に障害が発生していないかどうかを検査する (S3001)。

【0072】

障害が検出されない場合には (S3002)、CPU 112 は共有メモリ 120 のハートビートマークを更新する (S3003)。ハートビートマークは、クラスタ 180 内の各 CHN 110 が相互に動作状態を確認し合うための情報である。すなわち各 CHN 110 は、定期的に共有メモリ 120 の所定の領域にハートビートマークを書き込むことにより自己の処理が正常に行われていることを他の CHN 110 に示すと共に、他の CHN 110 のハートビートマークを確認する。これにより、クラスタ 180 内の CHN 110 の異常を相互に監視することができる。

【0073】

ハートビートマークには、CHN 110 の識別子や、CHN 110 が正常であるか異常であるかを示す符号、更新時刻等の情報が含まれる。続いて CPU 112 は、クラスタ内の他の CHN 110 のハートビートマークを共有メモリ 120 から読み出して、正常にハートビートマークが更新されているかどうかを確認する (S3004)。クラスタ内の他の全ての CHN 110 のハートビートマークが正常に更新されている場合には障害が発生していないと判断して (S3005)、S3000 からの処理を繰り返す。

【0074】

一方、S3004 で他の CHN 110 のハートビートマークを確認した結果、一定時間を経過しているにも拘わらず未更新であるものや、障害発生を示す符号が記載されているものを発見した場合には、CPU 112 はフェイルオーバ処理を開始する。まず CPU 112 は I/O プロセッサ 119 に、障害が発生した CHN 110 に対してリセット要求を送信させる (S3006)。障害が発生した CHN 110 から上記リセット信号を受領した旨のメッセージを受信したら、CPU 112 は、障害が発生した CHN 110 に関する引継情報を共有 LU 310 から取得

する (S3007)。

【0075】

共有LU310に格納される引継情報は図18に示す通り、ロック情報801、Samba712の構成情報802、マウント情報803である。ロック情報801は、障害が発生したCHN110が管理していたファイルロックテーブル721とLUロックテーブル722とを含む。マウント情報803は、障害が発生したCHN110が管理していたLUに構築されているファイルシステムのマウントに関する情報である。引継情報は上記以外にもネットワークインタフェース部111のMAC (Media Access Control) アドレスやIP (Internet Protocol) アドレス、ネットワークファイルシステムのエクスポート情報等があるが、これらは共有メモリ120から取得する。

【0076】

上記情報を取得したCPU112は、引継処理を行う (S3008)。まず障害が発生したCHN110のMACアドレスやIPアドレスを自己のネットワークインタフェース部111に設定する。これにより処理を引き継いだCHN110は、障害が発生したCHN110への情報処理装置1乃至3 (200) からのアクセスに応答することができるようになる。また障害が発生したCHN110のマウント情報に基づいて、障害が発生したCHN110のファイルシステムを自己のファイルシステムにマウントする。これにより処理を引き継いだCHN110は、それまで障害が発生したCHN110が管理していたLU310に対してアクセスすることができるようになる。この時、障害が発生したCHN110で管理されていたメタデータに異常がないかがチェックされる。メタデータの更新中に障害が発生する場合もあるからである。このチェックはメタデータを引き継いだCHN110のCPU112がメタデータチェックプログラムを実行することにより行われる。チェックの結果メタデータに異常が発見された場合には、メタデータの修復が行われる。メタデータの修正は、例えばUNIX (登録商標) 系オペレーティングシステムの場合には、iノードの管理情報が修正されることにより行われる。またCPU112は、ネットワークファイルシステムのエクスポート情報に基づいてファイルシステムをLAN400に接続された情報処理装置

1乃至3(200)に公開する。さらに、障害が発生したCHN110で実行中だった処理(チャネル制御部の処理)を引き継いで実行する。以上で引き継ぎ処理が終了する。

【0077】

一方、S3001で自己のCHN110に障害が発生していないかどうかを検査した結果、障害が発見された場合、あるいはS3000においてNASマネージャ706からフェイルオーバ指示があった場合には、CPU112は共有メモリ120のハートビートマークの更新を停止する(S3009)。なお障害の発生はS3001で行われる検査により検出されるのみならず、ハードウェア割り込みによっても検出される。その場合もCPU112は共有メモリ120のハートビートマークの更新を停止する。そうすると、クラスタ180内の他のCHN110によりハートビートマークの更新停止が検出され、フェイルオーバ処理が開始される。

【0078】

ハートビートマークの更新を停止したCHN110には、処理を引き継ぐCHN110のI/Oプロセッサ119からリセット要求が送信される(S3010)。このリセット要求は、処理を引き継ぐCHN110において実行される障害管理プログラム705処理ルーチンのS3006において送信されるものである。そうするとハートビートマークの更新を停止したCHN110は、上記リセット信号を受領した旨のメッセージを返信し、閉塞処理を開始する(S3011)。

閉塞処理はメモリ113のダンプを取得することにより行われる。メモリ113のダンプを取得するとは、メモリ113に記憶されているデータをLU310に記録することをいう。

【0079】

なお、3枚以上のCHN110でクラスタを構成するようにすることもできる。この場合は多段のフェイルオーバ処理を行うことが可能である。すなわち、フェイルオーバにより処理を引き継いだCHN110に障害が発生した場合には、さらに他のCHN110に処理を引き継がせるようにすることができる。この場合、最終的に処理を引き継ぐCHN110は、過去のフェイルオーバ処理において引き継がれてきた全ての処理を引き継ぐことになる。

【0080】

以上のフェイルオーバー制御がクラスタ180内のCHN110において行われる。本実施の形態に係るストレージシステム600においては、記憶デバイス制御装置100のスロット内にCHN110、CHF110、CHA110を混在させて装着することが可能であるが、このような構成の下においてクラスタ180を構成することができ、クラスタ180内ではCHN110に障害が発生しても、他のCHN110に処理を引き継がせるようにすることができる。

【0081】

===ファイルの高速アクセス===

次に本実施の形態に係るファイルの高速アクセス制御について説明する。本実施の形態に係るファイルの高速アクセス制御は、情報処理装置200から、記憶デバイス300に記憶されているファイルデータに対して、SAN500を介したブロック単位の高速なデータアクセスを行うための制御である。ファイルの高速アクセスを行うためには、情報処理装置200は、CHN110及びCHF110の両方と通信可能に接続されていることが必要である。図1においては情報処理装置3(200)が該当する。ファイルの高速アクセス制御を説明するためのブロック図を図19に示す。

【0082】

情報処理装置3(200)は、LAN400を介してCHN110と接続されている。また情報処理装置3(200)は、SAN500を介してCHF110とも接続されている。これにより情報処理装置3(200)は、CHN110を通じてLU310に記憶されているファイルデータをアクセスすることもでき、CHF110を通じて上記同一データをアクセスすることもできる。ただし、CHN110を介してアクセスする場合はファイル単位でのアクセスとなるのに対し、CHF110を介してアクセスする場合はブロック単位でのアクセスとなる。

【0083】

通常、情報処理装置3(200)がCHN110を介してLU310に記憶されているデータをアクセスする場合には、CHN110に対してファイル名を指

定したファイルアクセス要求を行うが、本実施の形態に係るファイルの高速アクセス制御によりLU310に記憶されているデータをアクセスする場合には、まずCHN110に対してファイル名を指定してメタデータ730（ファイルについての記憶装置の記憶領域上の記憶位置を特定する情報）の要求（リクエスト）を行う。メタデータ730の要求を受け付けたCHN110は、メモリ113またはキャッシュメモリ130に記憶されている当該ファイル名に対応するメタデータ730を読み出す。そして読み出したメタデータ730をLAN400を介して情報処理装置3（200）に送信する。なお、図20に示すようにメタデータ730はLU310にも記憶されているので、CHN110はLU310からメタデータ730を読み出すようにすることもできる。またCHN110は、情報処理装置3（200）からメタデータ730の要求を受信した際に、ファイルロックテーブル721のロック状態を確認することにより、当該ファイルデータについての排他制御を行うようにすることもできる。

【0084】

情報処理装置3（200）はメタデータ730を取得することにより、当該ファイルの記憶位置やデータサイズを知ることができる。情報処理装置3（200）はこれらの情報に基づいてファイルデータに対するブロックアクセス要求を生成できる。そして当該ブロックアクセス要求をSAN500を介してCHF110に対して送信する。

CHF110は、ネットワークインタフェース部111により上記ブロックアクセス要求を受け付ける。そしてI/Oプロセッサ119は当該データの記憶位置とデータ長等を抽出し、上記ブロックアクセス要求に対応するI/O要求を生成してディスク制御部140に出力する。このようにしてデータの読み出しや書き込み等が行われる。

SAN500はLAN400と比較して高速なデータ転送が可能なネットワークであるので、記憶デバイス300に記憶されているファイルデータに高速にアクセスすることができる。

【0085】

情報処理装置3（200）は、記憶デバイス300からファイルデータを読み

出す場合には、CHF110に対して当該ファイルデータのアドレスとサイズを指定してブロック単位でのデータ読み出し要求を送信する。CHF110は記憶デバイス300から読み出したデータをSAN500を介して情報処理装置3(200)に送信する。情報処理装置3(200)は、データをCHF110から取得したら読み出し処理を終了する。なお、CHN110からメタデータ730を取得する際に当該ファイルに対してロックを掛けていた場合には、CHN110に対してロック解除要求を送信する。

【0086】

一方、記憶デバイス300にファイルデータを書き込む場合には、情報処理装置3(200)は、当該書き込みデータと共に書き込みデータのアドレスとサイズを指定してブロック単位でのデータ書き込み要求をCHF110に対して送信する。CHF110は当該書き込みデータを記憶デバイス300に書き込み、書き込み完了メッセージを情報処理装置3(200)に送信する。情報処理装置3(200)はCHF110から書き込み完了のメッセージを受信したら、CHN110に対してメタデータ730の更新を要求する。なお、CHN110からメタデータ730を取得する際に当該ファイルに対してロックを掛けていた場合には、CHN110に対してロック解除要求も送信する。

【0087】

本実施の形態に係るファイルの高速アクセス制御は、データサイズの大きなファイルにアクセスする場合に効果大きい。データサイズの大きなファイルへのアクセスを高速なSAN500を介して行うことにより、ファイルデータに対する読み出し、書き込みの時間を短縮できる。これは、本実施の形態に係るストレージシステム600においては、記憶デバイス制御装置100のスロット内にCHN110、CHF110、CHA110を混在させて装着することが可能であり、CHN110を介したデータアクセスとCHF110を介したデータアクセスのそれぞれの特長をうまく利用することができることにより、実現できるのである。

【0088】

以上本実施の形態について説明したが、上記実施例は本発明の理解を容易にす

るためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に、本発明にはその等価物も含まれる。

【 0 0 8 9 】

【発明の効果】

記憶デバイス制御装置、及びデバイス制御装置の制御方法を提供することができる。

【図面の簡単な説明】

【図 1】 本実施の形態に係るストレージシステムの全体構成を示すブロック図である。

【図 2】 本実施の形態に係る管理端末の構成を示すブロック図である。

【図 3】 本実施の形態に係る物理ディスク管理テーブルを示す図である。

【図 4】 本実施の形態に係る LU 管理テーブルを示す図である。

【図 5】 本実施の形態に係るストレージシステムの外観構成を示す図である。

【図 6】 本実施の形態に係る記憶デバイス制御装置の外観構成を示す図である。

【図 7】 本実施の形態に係る CHN を示す図である。

【図 8】 本実施の形態に係る CHF、CHA を示す図である。

【図 9】 本実施の形態に係るメモリに記憶されるデータの内容を説明するための図である。

【図 1 0】 本実施の形態に係るディスク制御部を示す図である。

【図 1 1】 本実施の形態に係るソフトウェア構成図である。

【図 1 2】 本実施の形態に係るチャネル制御部においてクラスタが構成されている様子を示す図である。

【図 1 3】 本実施の形態に係るメタデータを示す図である。

【図 1 4】 本実施の形態に係るロックテーブルを示す図である。

【図 1 5】 本実施の形態に係るファイル単位での排他制御を説明するためのフローチャートである。

【図16】 本実施の形態に係るLU単位での排他制御を説明するためのフローチャートである。

【図17】 本実施の形態に係るフェイルオーバ制御を説明するためのフローチャートである。

【図18】 本実施の形態に係るフェイルオーバ制御を行うために共有LUに記憶されるデータの内容を説明するための図である。

【図19】 本実施の形態に係る高速ファイルアクセス制御を説明するためのブロック図である。

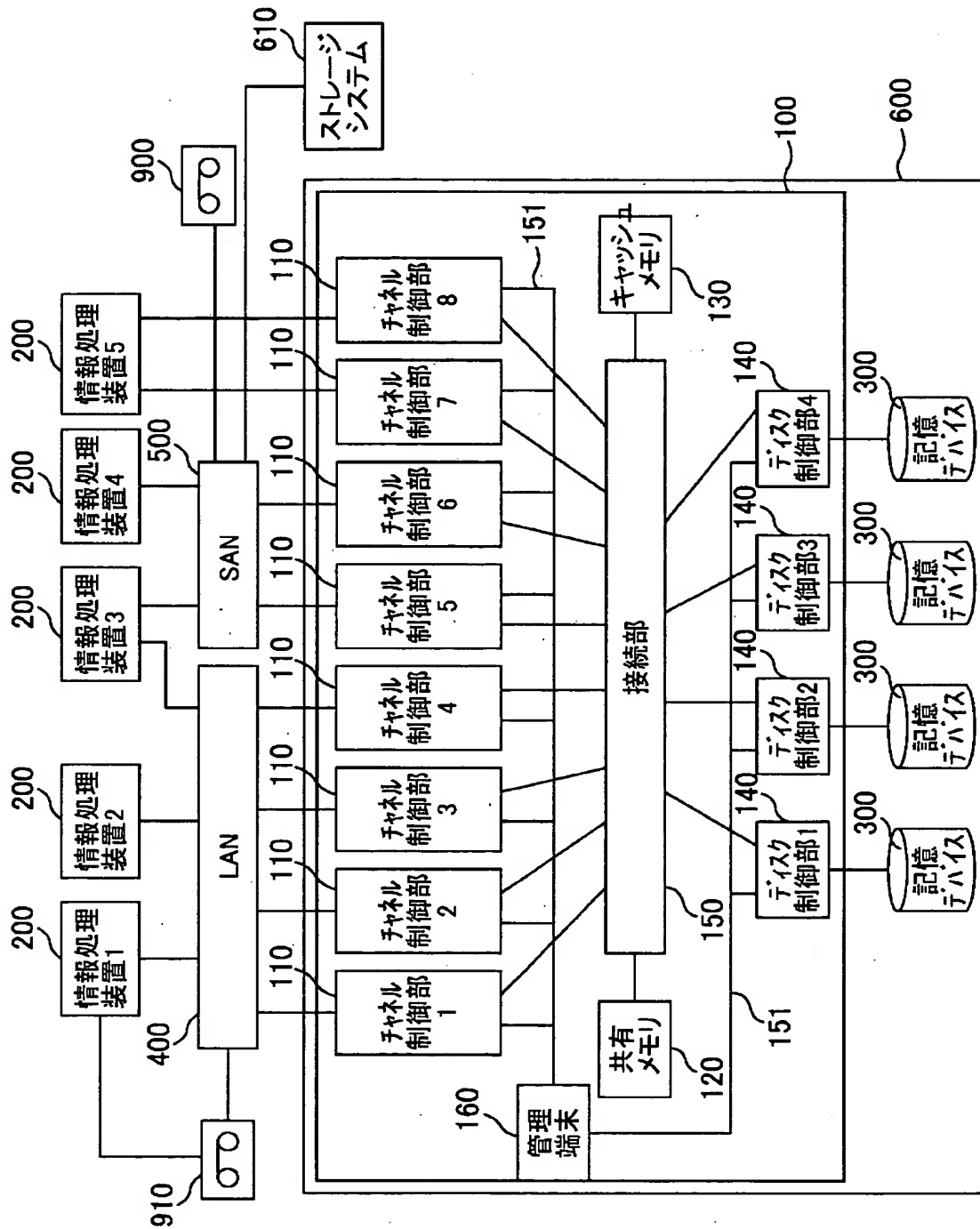
【図20】 本実施の形態に係る高速ファイルアクセス制御を行うためにLUに記憶されるデータの内容を説明するための図である。

【符号の説明】

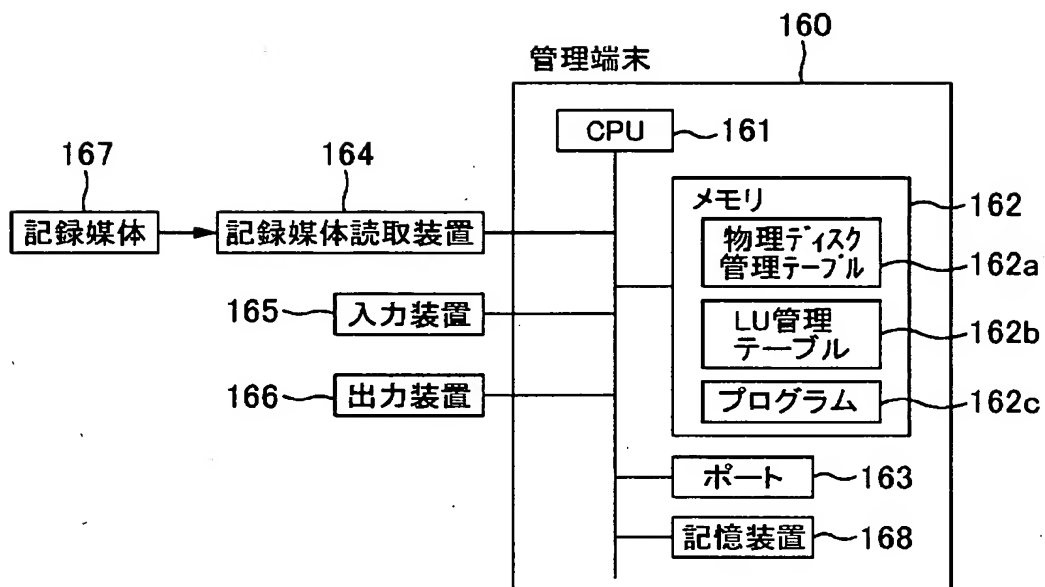
100	記憶デバイス制御装置	110	チャンネル制御部
111	ネットワークインタフェース部	112	CPU
113	メモリ	114	入出力制御部
115	NVRAM	116	ボード接続用コネクタ
117	通信コネクタ	118	回路基板
119	I/Oプロセッサ	120	共有メモリ
130	キャッシュメモリ	140	ディスク制御部
150	接続部	151	内部LAN
160	管理端末	170	ファン
180	クラスタ	200	情報処理装置
300	記憶デバイス	400	LAN
500	SAN	600	ストレージシステム
730	メタデータ	721	ファイルロックテーブル
722	LUロックテーブル		
900	SAN対応バックアップデバイス		
910	バックアップデバイス		

【書類名】 図面

【図 1】



【図 2】



【図 3】

162a 物理ディスク管理テーブル

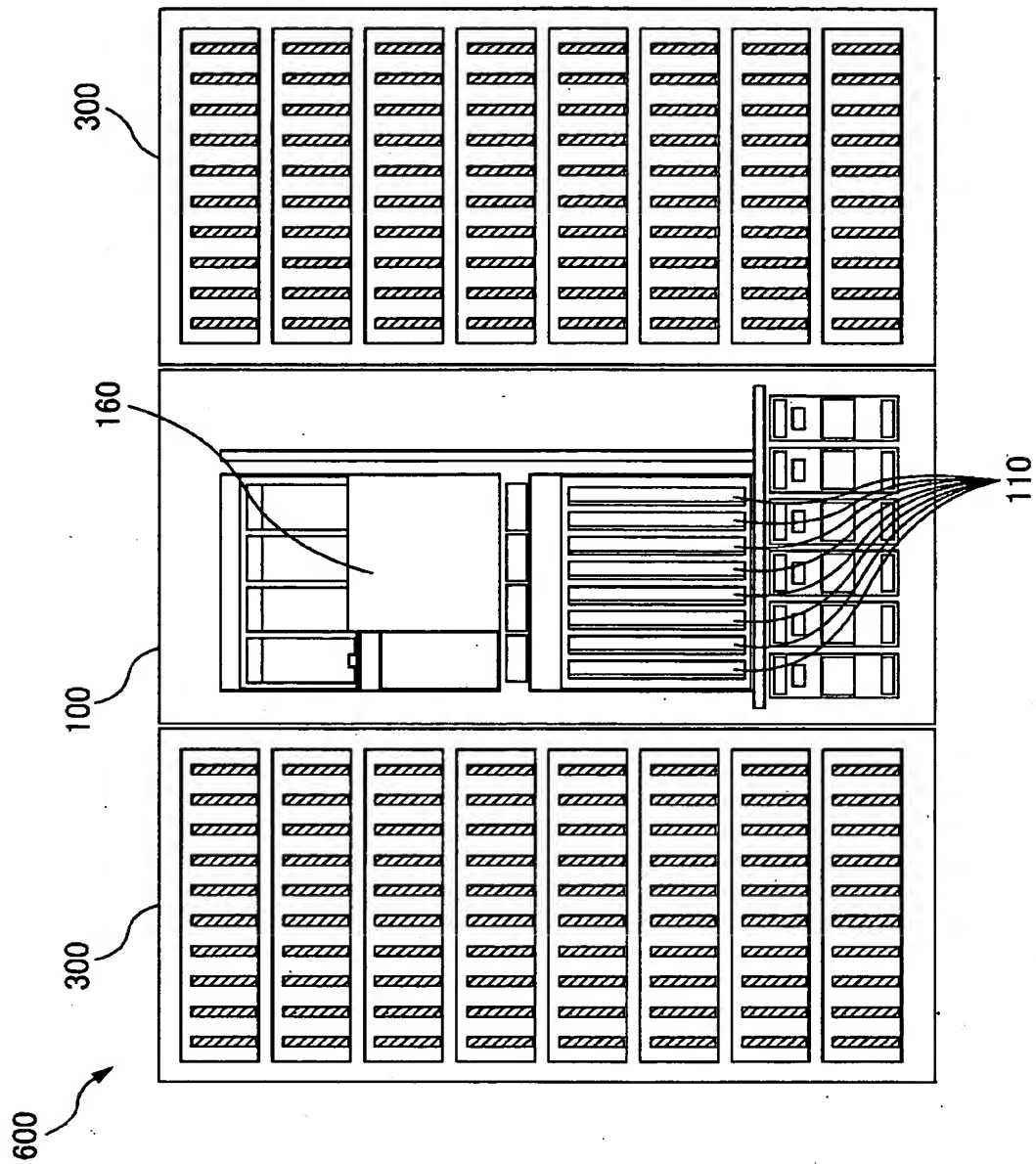
ディスク番号	容量	RAID	使用状況
#001	100GB	5	使用中
#002	100GB	5	使用中
#003	100GB	5	使用中
#004	100GB	5	使用中
#005	100GB	5	使用中
#006	50GB	—	未使用
⋮	⋮	⋮	⋮

【図 4】

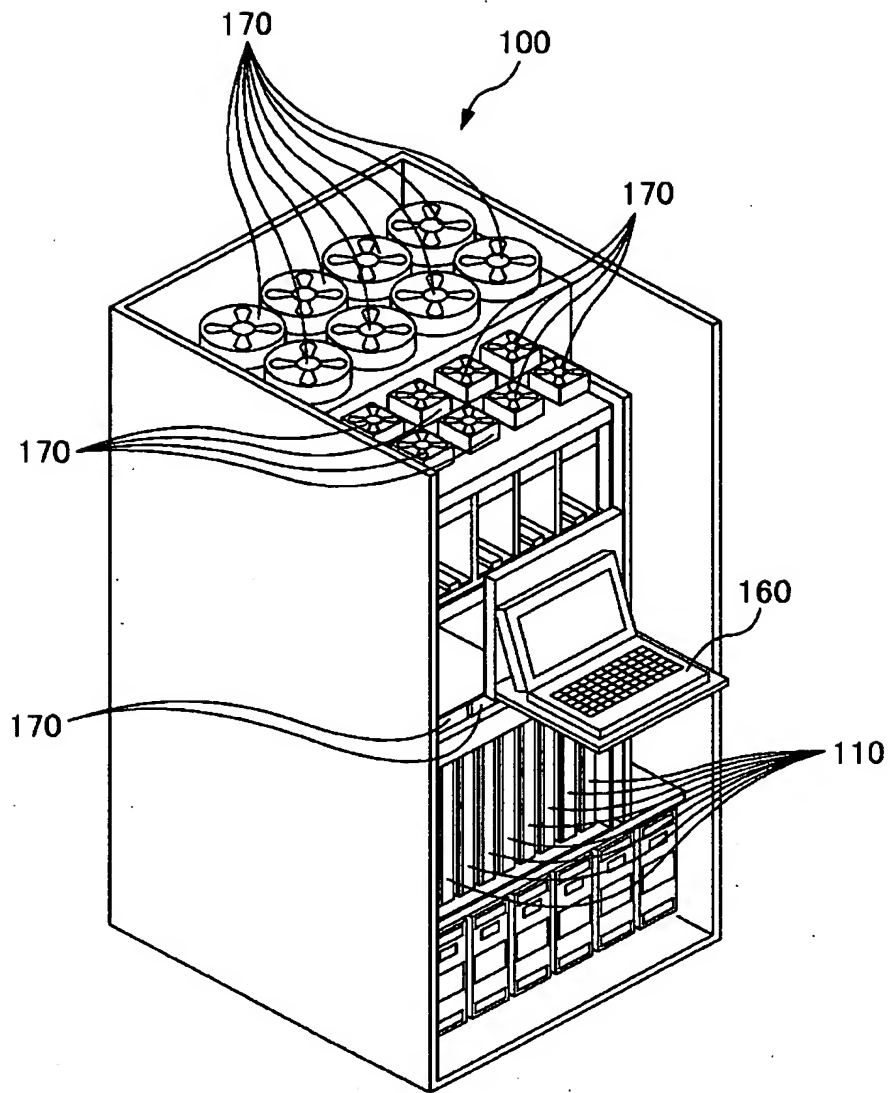
162b LU管理テーブル

LU番号	物理ディスク	容量	RAID
#1	#001,#002,#003,#004,#005	100GB	5
#2	#001,#002,#003,#004,#005	300GB	5
#3	#006,#007,	200GB	1
⋮	⋮	⋮	⋮

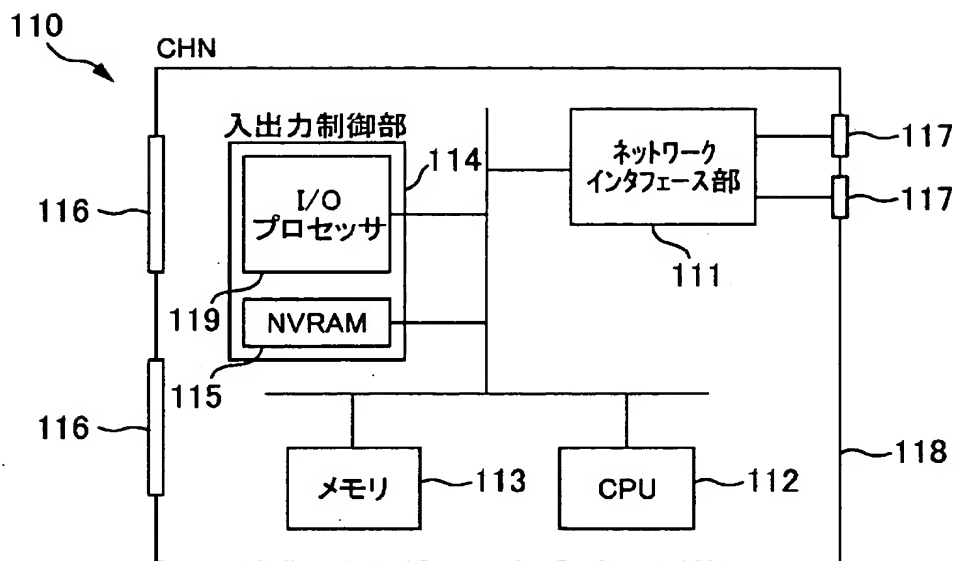
【図 5】



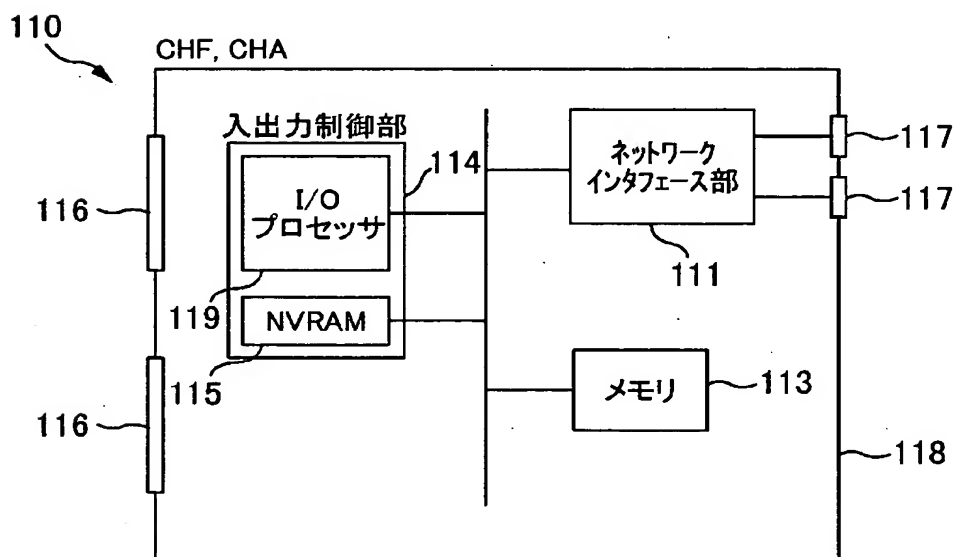
【図 6】



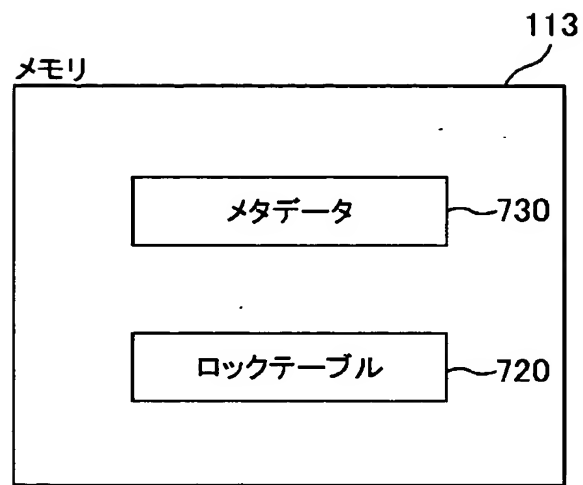
【図 7】



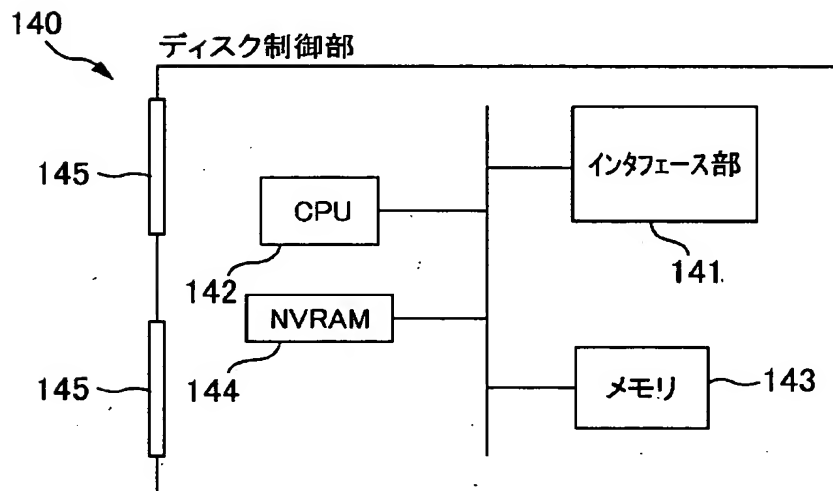
【図 8】



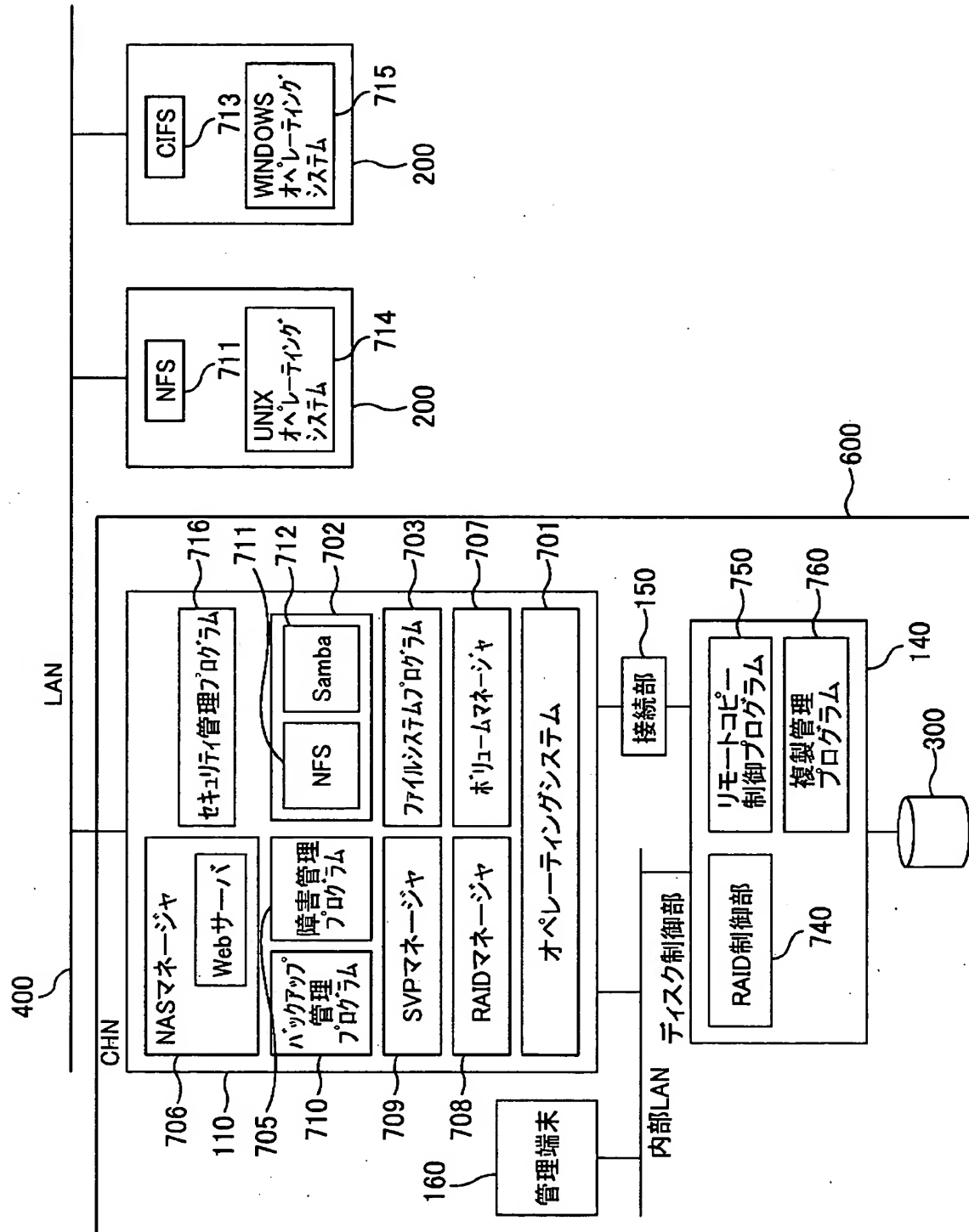
【図 9】



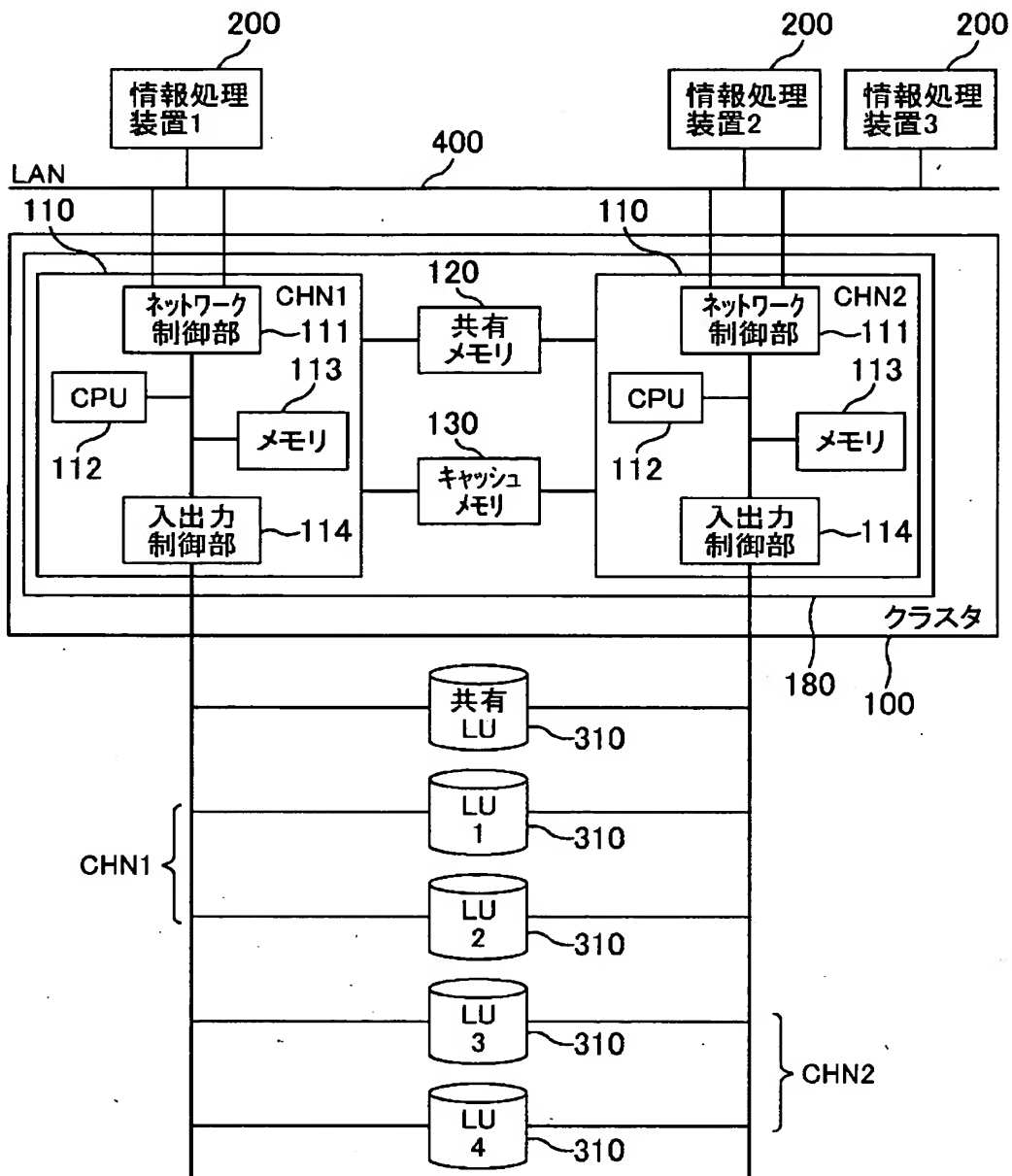
【図 10】



【図 11】



【図 12】



【図 1 3】

730

メタデータ

ファイル名	先頭アドレス	容量	所有者	更新時刻
A	7BSA	200MB	X	0:00
B	05BF	50MB	X	7:57
C	1F30	100MB	Y	9:15
D	470B	100MB	Z	15:20
⋮	⋮	⋮	⋮	⋮

【図 1 4】

721

ファイルロックテーブル

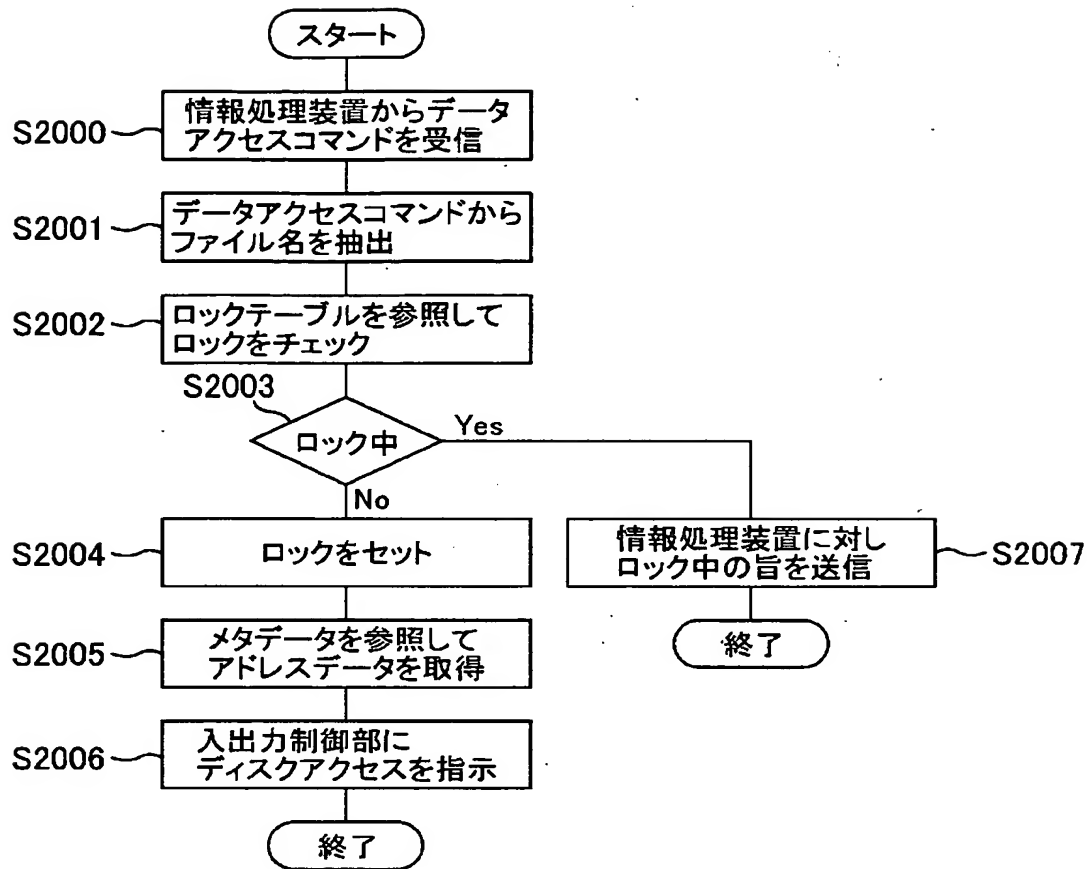
ファイル名	ロック状態
A	ロック中
B	—
C	—
D	ロック中
⋮	⋮

722

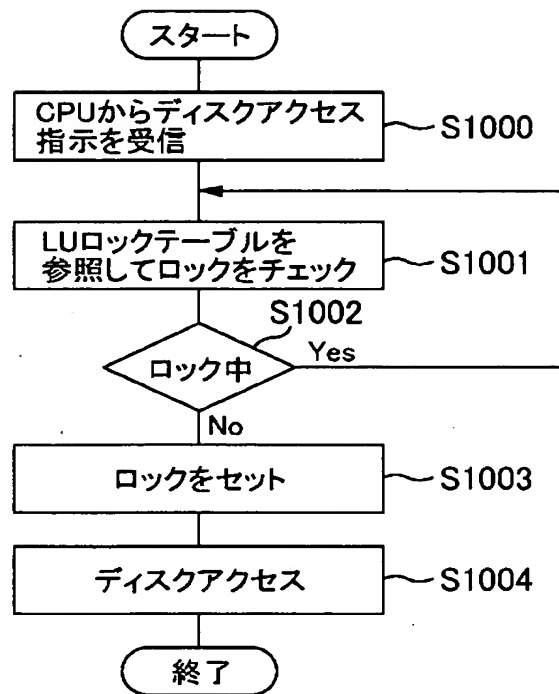
LUロックテーブル

LU	ロック状態
共有	—
1	ロック中
2	—
⋮	⋮

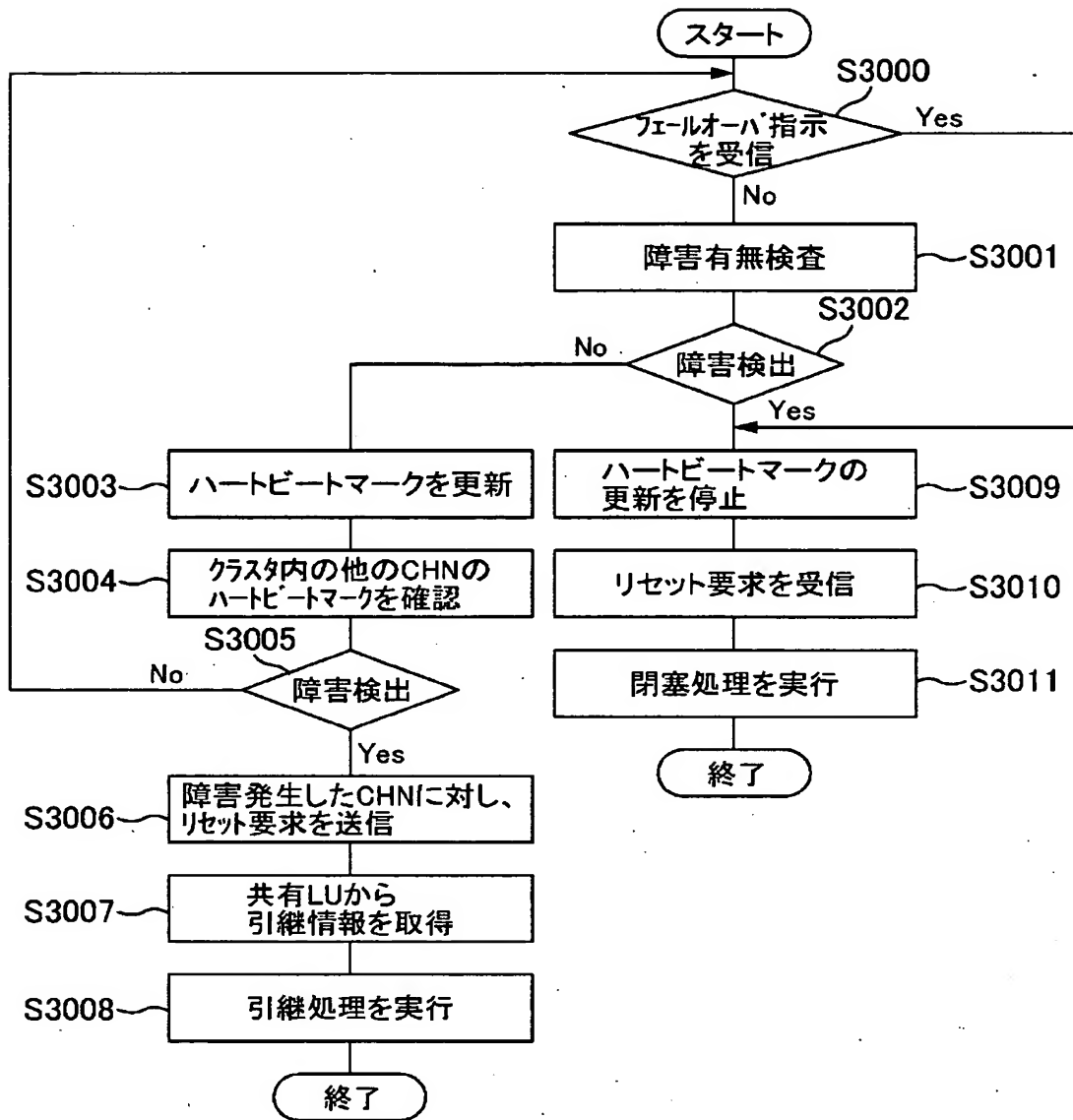
【図 15】



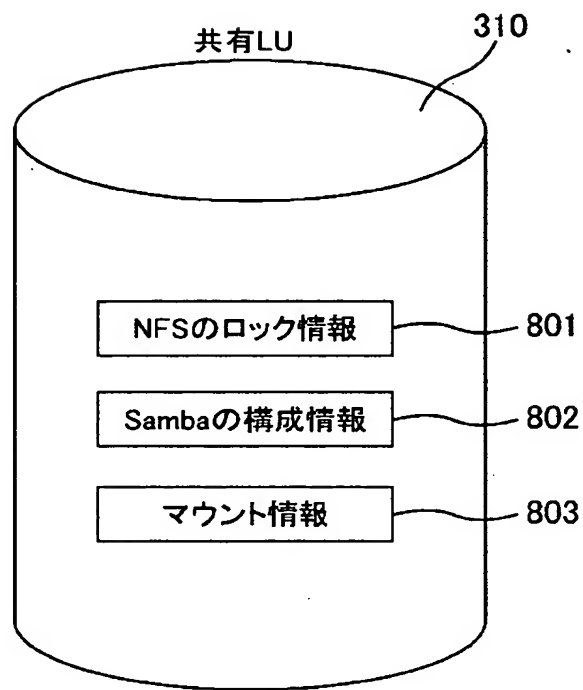
【図 16】



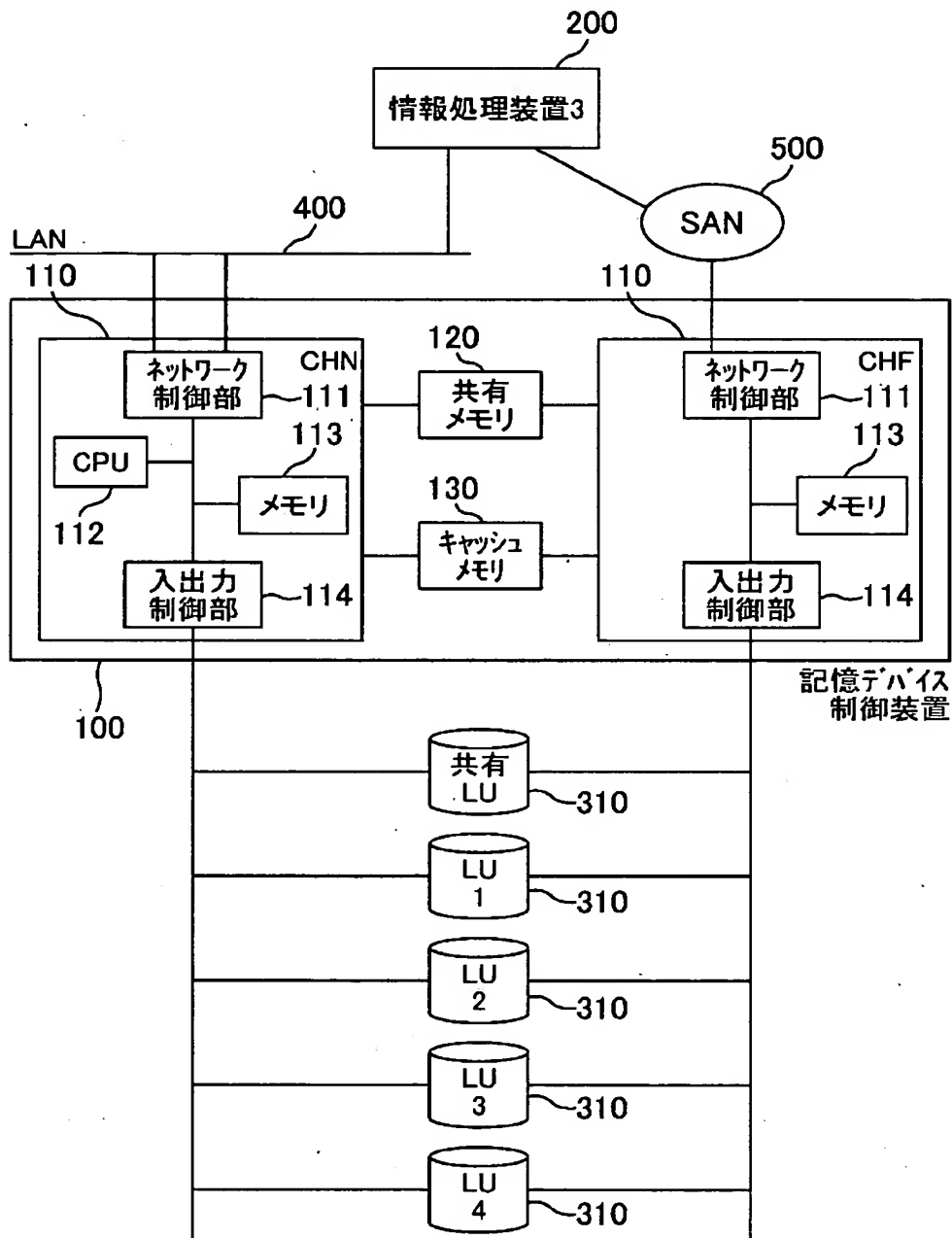
【図 17】



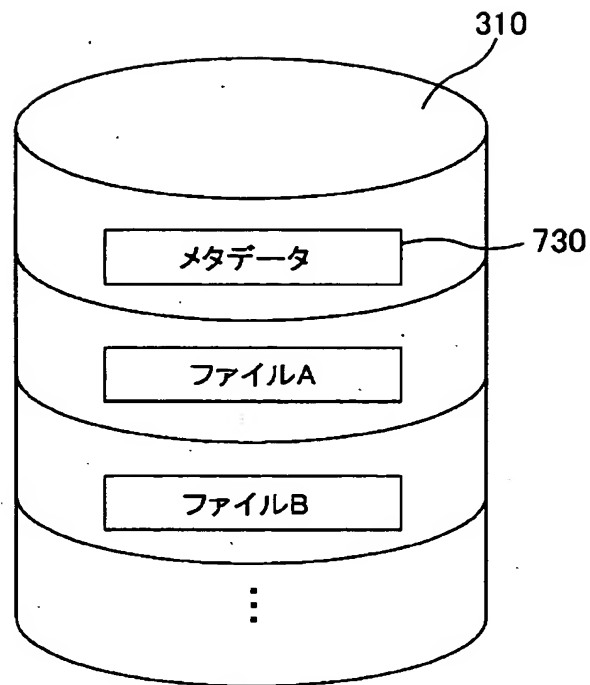
【図 1 8】



【図 19】



【図 2 0】



【書類名】 要約書

【要約】

【解決手段】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有するチャネル制御部を備える記憶デバイス制御装置であって、前記チャネル制御部が、あるファイルに対する前記データ入出力要求を前記情報処理装置から受信した場合に、そのファイルに対する排他制御を実行する排他制御手段を備える。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地
氏 名 株式会社日立製作所